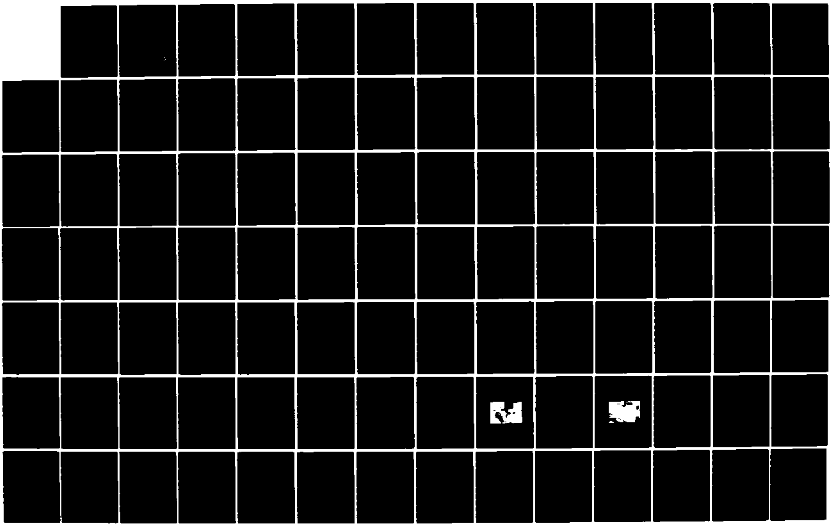
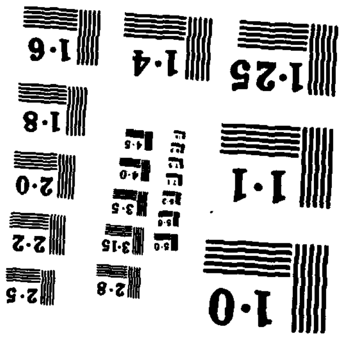
AD-A151 917 A COMPARISON OF PICTORIAL AND SPEECH WARNING MESSAGES 1/2
IN THE MODERN COCKPIT(U) AIR FORCE INST OF TECH
WRIGHT-PATTERSON AFB OH C P ROBINSON DEC 84
UNCLASSIFIED AFIT/CI/NR-85-23T F/G 9/3 NL





AD-A151 917

REPORT DOCUMENTATION PAGE — READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFIT/CI/NR 85-23T

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle): A Comparison Of Pictorial And Speech Warning Messages In The Modern Cockpit

5. TYPE OF REPORT & PERIOD COVERED: THESIS/~~DISSERTATION~~

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s): Christopher Paul Robinson

8. CONTRACT OR GRANT NUMBER(s)

9. PERFORMING ORGANIZATION NAME AND ADDRESS: AFIT STUDENT AT:

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS: AFIT/NR WPAFB OH 45433

12. REPORT DATE: Dec 1984

13. NUMBER OF PAGES: 147

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

15. SECURITY CLASS. (of this report): UNCLASS

15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report): APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES: APPROVED FOR PUBLIC RELEASE: IAW AFR 190-1

LYNN E. WOLAVER
Dean for Research and Professional Development
AFIT, Wright-Patterson AFB OH

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

DTIC ELECTE MAR 2 7 1985 S D E

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

ATTACHED

DTIC FILE COPY

DD FORM 1473, 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASS

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

85 03 11 043

ABSTRACT

Robinson, Christopher Paul, 1Lt, USAF. 158 pages. MSIE, Purdue University, December 1984. A Comparison of Pictorial and Speech Warning Messages in the Modern Cockpit. Major Professor: Ray E. Eberts.

A current trend in cockpit design ~~is~~ to incorporate synthesized speech to present secondary information to the pilot in an attempt to reduce mental workload, and to allow the pilot to keep his or her view out of the cockpit. Theories of multiple resource information processing support both of these reasons to use synthesized speech, but theories of stimulus - central processing - response (S-C-R) compatibility suggest the possibility that spatial information presented visually may have some distinct advantages over speech even though it uses the same input modality as the primary (flying) task. If the response is to be manual, then spatial information is more compatible ~~as~~ it can provide a direct mapping, or high ~~S-R compatibility~~ which can also reduce the ~~mental workload.~~ Twenty subjects participated in three dual-task experiments which compared tracking and emergency response performance when information was presented in the visual/spatial (pictorial) mode as opposed to the auditory/verbal (speech) mode. In all three experiments the pictorial mode elicited quicker response times, though in one experiment the pictorial mode also elicited more errors. Also, the pictorial subjects improved more with learning than did the speech subjects. While subjects were not successful at protecting their primary task when they added the secondary task, there were no interactions between task type and any other factor. These results indicate that more research concerning the spatial advantages of pictorial displays needs to be conducted before too many speech displays are incorporated into the cockpit.

A COMPARISON OF PICTORIAL AND SPEECH WARNING MESSAGES IN THE MODERN COCKPIT

A Thesis

Submitted to the Faculty

of

Purdue University

by

Christopher Paul Robinson

In Partial Fulfillment of the

Requirements for the Degree

of

Master of Science in Industrial Engineering

December 1984

ACKNOWLEDGMENTS

One of the most valuable insights gained from this endeavor was the need for help from other people, and how willing they can be to give it if one just asks. Very special thanks to Professor Ray E. Eberts, who spent many patient hours guiding my efforts with clear wisdom (even when I was too stubborn to listen with clear ears.) Thanks to Professor Gavriel Salvendy and Professor Robert D. Sorkin who also provided me with many good ideas of how to transform these efforts into a presentable document. Dr. John M. Reising of the U.S. Air Force Flight Dynamics Laboratory and Mr. David Jackson of the BDM Corporation both were very helpful during the experimental stages of this study; I am grateful to them. I am also very grateful to Mr. Stuart M. Weiss who was always so willing to help me with technical aspects such as statistical software and printout facilities. Finally, I cannot find enough words of thanks for my wife, Lt Cassandra S. Robinson, who kept me sane with her encouragement and her smiles throughout the program despite the hardships of our separate assignment locations.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Appendix
Figure

INTRODUCTION

Two new types of emergency information displays are currently being considered for implementation into aircraft cockpits: computer generated speech and computer generated pictorial displays. While both have advantages and disadvantages, basic theoretical as well as applied research studies have indicated that generated speech displays might have more advantages than pictorial displays. However, one of the primary advantages of pictorial displays, superior spatial coding, has generally been overlooked in those studies. Also, in previous comparisons between pictorial and speech displays the structure of the messages from the two groups has been different; and therefore has been confounded with the display type itself. The proposed research is an attempt to extricate the inherent spatial characteristics of pictorial displays; to study the possibility that when these are taken advantage of, pictorial messages are superior to speech warning messages.

In the "old control room", if the machine needed to inform the human operator of a problem, it could do so by

either flashing on a light or by sounding an auditory alarm such as a bell or buzzer. In this scenario, the human operator, once given the alarm, had to first decipher the alarm (e.g. distinguish it from the other alarms), then determine what to do about the problem, and finally respond to the problem. If the operator was lucky, the alarm sounder or light would be placed near the proper response control or at least near a display which he needed to attend to obtain more information about the problem. Such placement of the alarm signal helped direct the operator's attention to the proper area. In other words, the spatial location of the alarm helped decrease the operator's uncertainty of how to respond; it did some of the work by narrowing the operator's attention down to a specific section of the control console. Instead of the operator needing to decide which control out of a hundred to attend, he now only needs to decide which control out of ten requires his attention. But the control room would still be full of dedicated instruments, making it a formidable place for a human to enter, let alone operate efficiently during a high stress situation such as an emergency.

In the "new control rooms," however, the machine has available to it different, more versatile, methods of warning its human operator of impending danger. Instead of meters and dials each dedicated to a particular piece of

equipment, a cathode ray tube (CRT) can display pertinent information from any piece of equipment; instead of monitoring a hundred dials, the controller can monitor a few CRT's. Also, some flat-panel displays, such as plasma, thin film electroluminescent, liquid crystal, and side generated electron beam CRT's can now or will soon be able to replace dedicated dials. Additionally, instead of needing to memorize the meanings of forty different tones, buzzers and bells, the supervisor can listen to a synthesized speech message which tells him in his own language exactly where and what the problem is.

One particular type of control room which has been applying these new information presentation methods is the airplane cockpit. The proliferation of dedicated instruments in a cockpit, brought on by the advances in flight systems technology over the past three decades, has made this application very desirable. Not only are the number of subsystems increasing exponentially (Reising, 1975), but the space available in a cockpit is rather limited; it cannot easily be expanded to accommodate new instruments as can ground-based control rooms. Development of modern digital electronics has enabled a nearly simultaneous maturation of reliable computer graphics and speech synthesis with their potential application in the cockpit scenario.

Inevitably coupled with this growth is a certain competition between the methods. Which method should be used for displaying information regarding which subsystems? Some applications are clearly better suited for certain kinds of display methods, but other applications are not so clear. For example, a map of a strategic air strike area is clearly more effectively portrayed to the bomber pilot via graphical display than via digitized speech. On the other hand, it is not so clear whether an on-board system failure should be described to the pilot via graphics or via speech. There are advantages and disadvantages to both methods; these will be discussed later.

Currently synthesized speech is being considered in a number of these "unclear" areas for at least three reasons. First of all, as described by Butler, Manaker, and Obert-Thorn (1981), a primary goal of crew system engineers is to increase the time that the pilot can keep her head "out" of the cockpit so that visual contact with the target or the approaching runway is not interrupted more than absolutely necessary. At first glance, speech seems to facilitate this goal more than a visual display where the pilot must periodically bring the eyes back into the cockpit. A second reason, which will be discussed in further detail later, is certain information processing

theory which states that using a second input modality for the secondary information will incur less mental workload than using the same modality used for the primary (flying) task. Finally, speech synthesis may be receiving an undue amount of attention due to its novelty. There is an intrinsic excitement in being able to listen to a computer talk and being able to tell the computer, literally, what to do and have it respond. This advantage makes research and application of speech input/output easy to sell, while perhaps diverting some attention away from the potential advantages of visual/graphical information displays. Both speech and graphical displays have been incorporated into aircraft already; some examples will be included in the next section. As the designs of these information systems become further developed, however, their utilization will never become optimized without thorough individual research on both systems along with integrated research on possible combinations of the two methods.

Literature Review

Two levels of research have been conducted on generated speech and visual (CRT) display methods. The first level, applied research, has taken alternative methods of information presentation and compared them to each other. These methods include generated speech, auditory alarms (tones, horns, bells, etc.), pictorial, and alphanumeric

displays. The applied research studies are based on the other level, basic theoretical research. In this section, the current state of cockpit displays will be discussed, followed by a discussion of the applied research which has been conducted to improve the current state. Also, a review of the basic theoretical premises and models upon which applied research in information processing are based will be presented.

Current Cockpit Displays

Both commercial transport cockpits and military cockpits are now being equipped with synthesized speech and with CRT information displays. But the transition from conventional displays is gradual; the old electro-mechanical instruments and the buzzers and bells are still used extensively. In fact, while some of them are replaced by current - technology CRT's, the CRT display format is simply a close replication of the dial it replaced.

Kantowitz and Sorkin (1983) present a summary of auditory alerting methods used in a number of commercial aircraft cockpits. These include tones of various pitches, bells, whistles, wailers, chimes, horns, warblers, and speech. In some instances a cockpit alerting system includes up to forty different signals for alerting the pilots to the various possible problems or incoming

communication. Needless to say, this collection of signals imposes difficult training loads to say nothing of the memory requirements placed on the pilot. Miller (1956) discusses the limitations to the number of absolute pitches that a human can be expected to distinguish; the limit is about five or six tones. While the auditory warnings used in these cockpits include cues other than pitch, such as duration, repetition, and volume, their quantity still surpasses the recommendations of a number of documents (e.g. Cooper 1977).

CRT displays have found their place in commercial cockpits as is evidenced by their accepted use in the newest Boeing series, the 757 and 767. European airframers have also incorporated CRT displays in the Airbus A310 series (Reising, Emerson, and Aretz 1984). However, much of their use has been limited to alphanumeric printout, and the use of computer generated graphics has been limited to displaying updated "pictures" of the instruments that the CRT replaced. Instead of individual instruments displaying bank angle, false horizon, climb rate, and engine speed, for example, pictures of each of these instruments are drawn on the CRT. While this use of graphics appears somewhat unimaginative, Reising and Kopala (1982) point out that it may be a necessary transition from the conventional electro-mechanical instruments to acceptance of more

efficient, novel pictorial displays.

In the tactical military cockpit, both synthesized speech and CRT displays have been incorporated, though to a limited extent as in the commercial applications. In their study regarding the feasibility of implementing a synthesized speech warning system in the F-14, Butler et al. (1981) cite the current use of synthesized speech in the F-18 fighter and flight tests of a system in the F-15. As for the use of CRT displays, the implementation has been even more limited. The F-18, for example, uses CRT's to display information alphanumerically, but hardly any graphics are used.

So far, the imaginative use of available alternative methods for presenting secondary information to the pilot has been rather limited. This is due not only to the relative newness of these alternative systems, but also to the unresolved question of how to best utilize them. In the past few years there has been some research, though not a great deal, directed towards trying to improve these systems as well as to determine how and when they should be incorporated to elicit the quickest, most accurate, easiest pilot responses.

Applied Display Research

Two major categories of research, synthesized speech and spatial versus verbal methods, are dominating the field as alternative methods of information presentation. The motivation for synthesized speech comes largely from the current technological state of speech systems. Since speech input/output is not yet perfected, those involved in its development need to push to demonstrate its potential effectiveness; mostly by comparing speech systems to conventional warning buzzers and tones. The concept of pictorial presentation, however, has two "competitors": written text and spoken word, along with conventional displays. In discussing the recent literature covering research in application of visual and synthesized speech displays, it is convenient to break the field down into categories of generated speech displays and spatial versus verbal displays. Needless to say, the former category involves mostly the auditory modality, while the latter category involves both the auditory and the visual modalities. Theoretical bases for recent research as well as for the proposed study will be presented in the next section; the purpose of this section is to provide some examples of more empirical experiments which address the issue of how best to present secondary information to a system operator.

Werkowitz (1980) discusses some advantages and disadvantages to the application of speech generation systems in the cockpit. Some of these were referred to earlier. Advantages include: 1) speech messages can have an infinite set of messages as opposed to conventional warnings in which the pilot must memorize the meanings; 2) they are omnidirectional, such that the pilot need not look to find from what display the signal is emanating; 3) they can reduce visual workload; and 4) they can provide a good source of redundant information when coupled with visual displays. This last advantage was particularly evident in studies by Lilleboe (1963) and Stroface and Stark (1963) which used speech systems in conjunction with conventional warnings in actual in-flight tests. Potential disadvantages include: 1) interference with radio communications; 2) interference from cockpit noise; and 3) inability to convey information spatially.

Wicker (1980) discusses an experiment and an opinion survey regarding a cockpit speech interactive system. In the experiment which implemented such a system in a flight simulator, it was found that speech was indeed helpful to the pilots, but mostly so when it was used to reinforce visual displays. This corresponds to the "redundancy" advantage listed above. An interesting result of the

opinion survey, however, was that pilots felt that emergency systems ought not be actuated by speech input. As will be seen later, this suggests that for optimum compatibility the corresponding emergency warning messages also ought not be presented by speech output.

Mountford, North, Metz, and Warner (1982) ran an experiment which compared speech and manual input of navigation data in a dual task (tracking and data entry) situation. They found that, in the speech entry mode, less tracking error was incurred than in the manual mode. However, response time (time to complete data entry) did not significantly differ between the modes. While that study concentrated on response modes as opposed to perception modes, the results are considered applicable because of the consistency between input and output modalities.

Though not in a cockpit, Bouis, Voss, Geiser, and Haller (1979) studied various presentation methods of secondary information in an automobile. Methods included visual (text and lights), auditory (speech and tones), and combinations of these for binary and for multiple state (analog) information. Their measurements included tracking degradation, information processing (response time and intelligibility), and subjective preferences. The recommendations based on their results included 1) for

frequent binary alarms, use visual signals; 2) for rare binary critical alarms, use "dynamic sound" and lamp; 3) for textual information with many words, use speech with a preparatory signal; and 4) for road guidance, use pictorial presentation. Relating this to the cockpit suggests that for emergency warnings pertaining to main aircraft systems (rare binary), conventional tones and flashing lights might be best but for subsystems that require more detailed information presentation the speech system might be better.

A number of studies have been conducted on how speech synthesis, assuming its availability, should be implemented in the cockpit. Some results (Simpson, 1976; Simpson and Navarro 1984) apply to comprehension and intelligibility of the messages themselves. For example, if monosyllabic words are used in the message, they should include sentence context, but if polysyllabic words are used, sentence context is not necessarily advantageous for message comprehension (though it does improve response time from end of message). Other factors influencing the intelligibility of speech messages include speech rate, hardware used, and type of speech. Messages spoken at 156 words per minute were better than when spoken at 123 or 178 words per minute. Synthesized speech is more easily understood than digitized speech, and the male digitized speech is more

distinguishable in cockpit noise than digitized female speech. Finally, flight experience seems to have no effect on the intelligibility of digitized words.

Two further studies took a closer look at the advantages of semantic context, and the integration of speech with conventional tones. Simpson and Williams (1980) contend that with speech, critical flight information can be transmitted to the pilot without the pilot being distracted from visual tasks, especially VFR flying. Adding an extra word to the speech messages, while naturally lengthening the message delivery time, did not increase the reaction time from onset of message. This suggests that the extra semantic context actually reduced the pilot's mental workload as the time from end of message to response was significantly reduced. Unfortunately no mention is made of primary task degradation; one would suspect that less degradation might occur when the semantic context was provided. The other issue referred to was that of placing a warning tone before the speech message. With the alerting tone, the overall response time was increased, but not by the full amount of time allotted to the tone and pause. In other words, the response time from end of message was actually shorter than without the tone. Again this hints at a possible further reduction of workload. However, the authors concluded that the overall increase in response time

was more important than the decrease in workload. A measure on the primary task may have shed some different light on this matter. In the study, all messages were of the same nature: emergency warnings.

Hakkinien and Williges (1982), as referenced in Simpson and Navarro (1984) took a further look at the question of the tone preceding the speech message. In their study, speech messages were used for non-warning messages as well as warning messages. This study found that when a tone preceded the warning messages only, the response times were indeed reduced. This suggests that the tone acted similar to the semantic context of the Simpson and Williams (1980) study; it provided another level of context which reduced response time and workload.

This research exhibits the fact that speech input/output does indeed have a useful and advantageous place in certain aspects of pilot - computer interaction. They also show the level at which the research is being conducted; it is already at the point of trying to optimize the messages which will be given to the pilot. Spatial communication, on the other hand, appears to still be at the point of finding a proper niche in the cockpit, but not yet to the point of optimizing the views to be displayed.

While much of the research devoted to the application of pictorial information displays does not appear quite as in-depth as that devoted to synthesized speech as sampled above, there have been a number of studies comparing this method to speech and to alphanumeric CRT displays. Some research has been carried out which compares variations in pictorial format such as color versus black and white, and stroke (line drawings) versus "color raster" (filled-in drawings). However, little research has been completed in a theoretical optimization of pictorial displays. As stated above, the main thrust of pictorial display research has been as a comparison between spatial and verbal information presentation.

Hawkins, Reising, Lizza, and Beachy (1983) conducted a study comparing pictorial presentation of emergency information with text and speech while the subjects "flew" a combat mission in a simulator. Hypothesizing that the pictorial displays would be more effective than both alphanumeric and speech displays, the authors measured performance via horizontal and vertical tracking error, and via "task completion time." It should be noted that the eighteen subjects consisted of Air Force pilots and weapons systems officers who had all had training in the emergencies simulated in this paradigm. Also, the experimental design (repeated Latin Square) was selected to cancel a learning

effect. No significant effects were found among any of the three performance measurements. Results of a questionnaire, however, found a significant preference for the speech mode over both pictorial and alphanumeric but no difference between the latter two. The authors suggest that a possible reason for these results was that the subjects were familiar with the emergencies as described by the text and speech, but they were not familiar with the pictorial representations and therefore had to include an extra translation step in processing information presented in that mode. While many might argue that it is important to have actual pilots in this type of study, using non-pilots as subjects might alleviate this bias while allowing a purer test of the theoretical principles of interest. Analysis of the effects of practice and its interaction with treatments could also be an important factor which could not be determined in the present experimental design. This might have helped to isolate the experience factor; one might expect that while the pictures were more difficult to understand at first, the interaction between mode and practice would show greater improvement with the pictures than with text or speech. In any event, the speech mode did not outperform the spatial pictorial mode.

Williamson and Curry (1984) describe a dual task study aimed in part at comparing subjects' abilities to process

and report information which is presented vocally, textually, and pictorially. While "flying" a simulator (the flying task consisted of a videogame simulating a military attack mission), subjects were given information regarding fuel status, weapons status, or engine status in one of the three modes listed above. The secondary task consisted of retrieving and entering this information, either vocally or manually, into the on-board computer. In this experiment the subjects were college students, which helped to relieve the bias discussed in the Hawkins et al. (1983) study. Under the hypothesis that the flying task would be degraded less with the speech input and output conditions than with text, pictures, and manual responses, the authors actually found no significant differences in the flying task performance. A possible explanation for this is that subjects considered the flying task to be the "primary task". Thus, changing the difficulty of the secondary task (assuming the different modes incurred different difficulty levels), should not effect the primary task (Navon and Gopher, 1979), but should affect the secondary task performance. Indeed, significant performance differences were found in the secondary task. Analysis of the data entry showed that manual responses were initiated more quickly than vocal responses. Task completion times were correlated to the mode of information presentation; with textual and speech modes both eliciting shorter completion

times than the pictorial mode. No differences were found between the speech and text modes.

Although spatial information presentation was found to be worse than speech or textual, two elements of the study ought to be considered. First, the various system status displays contained up to four or five logical lines of information. A single picture containing all this information may have been too cluttered; two simpler pictures displayed consecutively may be a better method of presentation. Another possible bias may have been introduced by a disparity between the constructs of the vocal and the pictorial messages. For example, on the engine status pictorial display, five parameters are shown even though only two are seen to be out of tolerance. In the corresponding speech message, only the two parameters which are out of range are referred to. Greater parity, and thus a fairer comparison, might be achieved if the pictorial display only included those two parameters which were out of tolerance.

The second element of the Williamson and Curry (1984) study to be reconsidered is the spatial compatibility between the pictorial displays and the response buttons. To demonstrate the possible advantages of pictorial displays, this spatial relationship must be capitalized upon. In this

experiment the responses were not spatially formatted, therefore the subjects had to translate the pictorial spatial information into serial information before searching for the correct response. The indirect mapping of the pictures with the responses may have hurt the spatial mode in this spatial/verbal comparison.

In a study limited to the visual input modality, Aretz and Calhoun (1982) designed an experiment which compares different aspects of pictorial and alphanumeric displays and their integration with each other. In a fixed base simulator, subjects were required to maintain flight control while retrieving weapons stores information. This information was presented in four modes: 1) alphanumeric, 2) color pictorial, 3) black and white pictorial, and 4) a combination of alphanumerics and color pictorial. The subjects were experienced Air National Guard A-7 pilots. Results of this experiment indicated that alphanumeric displays had a shorter task completion time (information retrieval and response). However this method was not statistically better than the color pictorial or combination methods. The black and white pictures were significantly worse than each of the other three methods.

One questionable aspect of the Aretz and Calhoun (1982) study is the type of information which was presented for

retrieval. Many of the retrieval questions were quantitatively oriented, for example requesting the "number of stores selected." When the response is to be quantitative (or, verbal as opposed to a direct spatial translation), then greater compatibility is achieved when the information is presented in the verbal format as opposed to a spatial format. While some of the questions were spatially oriented, for example "type of fuzing selected," no interaction is reported between type of question and presentation method. One might expect that pictorial presentation would elicit quicker responses to "spatial" questions, while alphanumeric presentation would elicit quicker responses to "quantitative" questions. A different experimental design could facilitate the probe of this interaction.

The results of these experiments indicate that neither presentation mode, spatial or verbal, were clearly better than the other. Subjects' responses to a questionnaire demonstrated a preference for the color pictorial/alphanumeric combination displays. Another interesting aspect of the experiment is a potential contribution to the optimization of spatial displays discussed previously. Spatial displays should be formatted in color, not black and white.

Digressing briefly from the cockpit scenario, though not from the spatial/verbal question, an experiment by Tullis (1981) took a close look at the application of this question to trouble-shooting in a telephone system. Subjects were required to interpret the results of a telephone line test. These results were presented to the subjects in a variety of formats: 1a) alphanumeric - structured, 1b) alphanumeric - narrative, 2a) spatial - color, and 2b) spatial - black and white. The results of this experiment showed that response time was significantly shorter with spatial information than with alphanumeric - narrative information. However, after substantial practice, the alphanumeric - structured format induced nearly the same response times as the spatial displays. No differences in response accuracy were noted. While no significant differences were noted this time between performance with color and with black and white spatial displays, questionnaire responses did indicate a strong preference for the color spatial display.

Moroze and Koonce (1983) describe an experiment which tested differences, in a small fixed-base simulator, between traditional round-dial displays, digital Heads Up Displays (HUDs), and spatial HUD displays. The digital HUD displayed alphanumeric information while the spatial HUD incorporated linear tape indicators. The hypothesis of the study was

that the linear tape method would induce better performance than the other two methods because it provided information in a way that was more consistent with the subjects' internal mental models of the information. Without this consistency, the subject has to go through more coding transformation processes, which increases the probability of error as well as increasing the mental workload.

Another factor which must be considered in the experiment is the difference between "conventional" in-cockpit displays and HUD displays. The HUD display may be the closest that a visual display can come to satisfying the desires expressed by the "eyes out of cockpit" proponents of speech displays. Especially in the case of a spatially formatted HUD display, the pilot could use peripheral vision to gather pertinent information from the HUD while keeping the foveal vision fixed on the outside runway or target. Arguments have been presented that even with a HUD, focal considerations negate its usefulness. These arguments state that changing the eye's focal length from infinity to few feet (the distance to the HUD) take the same toll as diverting the eyes from an outside target to an inside instrument. On the contrary, spatial displays do not require the fine focus required by digital displays. From this viewpoint the spatial HUD display seems more attractive.

Returning to the Moroze and Koonce experiment, subjects were instructed to perform flight maneuvers while responding to a recognition test. This test consisted of picking out odd-even-odd sequences in a string of random digits presented auditorially. The only significant result obtained was on the run where performance criterion was met; the traditional round-dial display brought on better performance than the other two. After this run, no significant differences showed up. This could have been due to the tasks being too difficult or too easy. Another possibility is that the spatial display was not designed well; thus instead of outperforming the verbal display it merely matched the verbal display performance.

The cockpit studies mentioned thus far have included dual-task paradigms. An interesting study by Hartzell, Dunbar, Beveridge, and Cortilla (1983) involved a single task experiment meant to challenge tradition in the configuration of helicopter cockpits. Traditionally, the airspeed and altitude indicators have been arranged contralaterally with the corresponding controls. This means that the altitude indicator is located to the right of center panel and the airspeed indicator is located to the left. But the altitude control is operated by the left hand and the airspeed control is operated by the right. This,

they contended, introduced an incompatibility which caused poorer performance of the flight task than if an ipsilateral arrangement was incorporated. In the experiment, subjects had to maneuver the helicopter to a predetermined goal flight state which was represented on the altitude and airspeed displays. The results were as predicted; subjects consistently accomplished the tasks more quickly when the displays and controls were arranged ipsilaterally than when they were arranged contralaterally.

The relationship of the Hartzell et al. study to the question of spatial versus verbal displays is somewhat subtle. The main point is that if a spatial arrangement is to be used, it must make the most of its available information. In other words, a big advantage of a spatial display is that it can, more effectively than a verbal display, direct the observer to a correct manual response. But this advantage only holds if the spatial display is designed in a fashion which is compatible with the physical environment to which it refers.

The final study (Mazza, 1977) to be reviewed here perhaps unwittingly demonstrated an advantage of a spatial display. The effort was completed during relatively early stages of CRT application in military cockpits. The author was questioning the incorporation of CRT's basically because

they did not provide the spatial information which was inherent in conventional displays. For example, in a conventional arrangement, each engine has its own dedicated fire warning display. The location of this display corresponds closely to the proper response, i.e., the extinguisher for that engine. Presumably, a CRT display could not provide this inherent information.

The experiment compared a conventional warning display to an integrated display. The integrated display was meant to alleviate the overabundance of dedicated displays in modern cockpits caused by the proliferation of subsystems. In other words, one integrated display can take the place of a number of dedicated displays; the appropriate information being displayed only when needed. In the integrated display condition, the warning messages appeared alphanumerically on the CRT in the format, "ENGINE FIRE NO.1." All the possible messages appeared in the same location on the CRT when the particular emergency arose. The response panel was arranged as a two dimensional four by four keyboard. The left column was for engine number one, the second for number two, and the third for number three. The top row was for fire, the second row for oil pressure, and the third row for temperature. This arrangement corresponded nearly precisely with the way that engine warning lights are arranged in a conventional cockpit, and with the "conventional" display

condition used in the experiment. The other buttons were used for miscellaneous warnings.

As could be expected, the conventional display outperformed the integrated display. When a warning flashed up on the conventional display, the subject did not even need to read the lighted message; its spatial location could be directly mapped onto the response keyboard. With the integrated display, on the other hand, the message had to be read and then translated to a correct spatial mapping before the response could be made. Even though in the conventional displays the warning lamps included alphanumeric text (verbal information), it was most probably the spatial characteristics of these displays that caused their highly significant improvement in performance over the purely verbal information provided by the "integrated" CRT displays.

Mazza (1977) warned that changing from conventional displays to integrated displays would result in a loss of this clearly important spatial information. This would be an obvious negative aspect in the movement for fewer cockpit instruments. Modern computer graphics on a CRT, however, allow the possibility for both. On the one hand, integrated display systems such as CRT's can vastly reduce the number of instruments in the cockpit. On the other hand, they can

still provide all, and more, of the critical spatial information if they are designed with this in mind.

The experiments discussed up to this point have illustrated the type of research currently being conducted in the continuing effort to make the increasingly difficult tasks of today's pilots within the limits of human capabilities. With the concurrent development of synthesized speech systems and cockpit-compatible graphics systems, there has been a tendency toward designing empirical studies pitting speech input/output with visual and manual input/output. As occurs in all studies, there have been important factors, or limitations, in these studies which may have introduced certain biases in the results. A few of these factors have been mentioned already, for instance the optimization of the pictorial displays. Many of the pictorial displays used have not taken advantage of the basic benefits which can be derived from spatial information output. The directive compatibility with response, eliminating the need for verbal to spatial translation, is one of these benefits.

Many of the previous experiments have tested trained pilots. Trained pilots have developed stereotypes as to how information is, and ought to be, displayed. These stereotypes can interfere with the subjects' response

performance when information is presented in a different manner from what they are used to. Many of the pilots used in the studies have already been exposed to speech displays in the cockpit; none have been exposed to an extensive use of pictorial graphics. The time limits imposed on experiments do not allow for a comprehensive training period which would help eliminate the stereotype bias. Therefore subjects with little training in either display mode; i.e., non-pilots, provide a better control. The use of non-trained pilots has been argued as defeating the inference space in which we are interested: trained pilots. But, we are not only interested in empirical studies which will determine what display type to install in all cockpits this minute. We are interested in theoretically based concepts of information processing which apply to the human mind in general, and which will direct the application of systems into future cockpits and future training methods.

The review of literature so far has concentrated on experiments which, though based on theoretical premises, have been somewhat empirical in nature. Starting with a theoretical background, the studies have narrowed down the application inference to the pilot-cockpit interface. The next section of this paper will review some of these major theoretical premises as they apply to human performance in general.

Basic Research

There are four basic theoretical premises describing human information processing which are applicable to a high workload situation during which quick and accurate responses are required. These include the theories of multiple resources and stimulus-central processing-response (S-C-R) compatibility. They also include the concepts of mental models and of hierarchical mental organization. This section will take a look at these theoretical viewpoints and concepts as background for the experiments which were conducted under this effort.

Navon and Gopher (1979) present a comprehensive overview and the implications of a multiple resource theory of human information processing. This model merges and expands upon the previous processing theories of single capacity (Kahneman, 1973) and multiple channels (Allport, Antonis, and Reynolds, 1972). Through this merging of theories, certain identified limitations of each are bridged and explained by the combination.

The essence of the multiple resource theory is that the information processing system consists of a number of pools, from which resources can be drawn and allocated to a set of processes simultaneously. Each pool has its own capacity, or limit, of resources. This is not to say that each pool

can only be devoted to one task as suggested by a multiple channel theory. Rather, if two tasks are being processed simultaneously, both tasks may draw resources from the same pool though the total amount of resources allocated can not exceed the capacity of that pool.

In a single capacity model the brain is considered to have one central pool of resources (and its corresponding limit) from which simultaneous processes compete for allocation of the resources. As Navon and Gopher (1979) state, a limit to this notion is demonstrated when "the performance of a certain task is disrupted more than the performance of another one by pairing either of them with a third one, [but is] disrupted less by a fourth one." (p. 232) The difference, then, that the multiple resource model provides is that there are a number of resource pools, each with their own capacities. When two or more tasks are performed simultaneously, it is an interaction of multiple capacity limits which determines the performance rather than one central limit.

In previous multiple channel models, it was theorized that information is processed through a number of channels but each of these channels could only handle one process at a time. Again Navon and Gopher point out a problem with this line of thought: "...[their] model seems inadequate

once we realize that processes that use the same mechanisms sometimes interfere with each other but seldom block each other completely." (p. 233) The multiple resource concept allows for this contingency in that each channel may actually support more than one task at a given time. Likewise, the tasks are accomplished by drawing from a combination of the various resource pools, not just one channel.

Wickens' (1980,1984) multiple-resource (see Figure 1) model breaks the resource pools down into divisions of stages, modalities, codes, and responses, and shows the relationship of each to the others. The stages are divided into two main processes: 1) encoding and central processing, and 2) responding. The first process includes the perception and mental processing of information, while the second process is the physical response. The modality categorizes the encoding mechanism; by eye (visual), or by ear (auditory). Two different types of information can be received: spatial and verbal, which correspond roughly to analog and digital information. Finally, responses can be made manually or vocally.

The model has implications for both single and dual task performance. Regarding single task performance, Figure 1 suggests that if information is encoded and processed in a

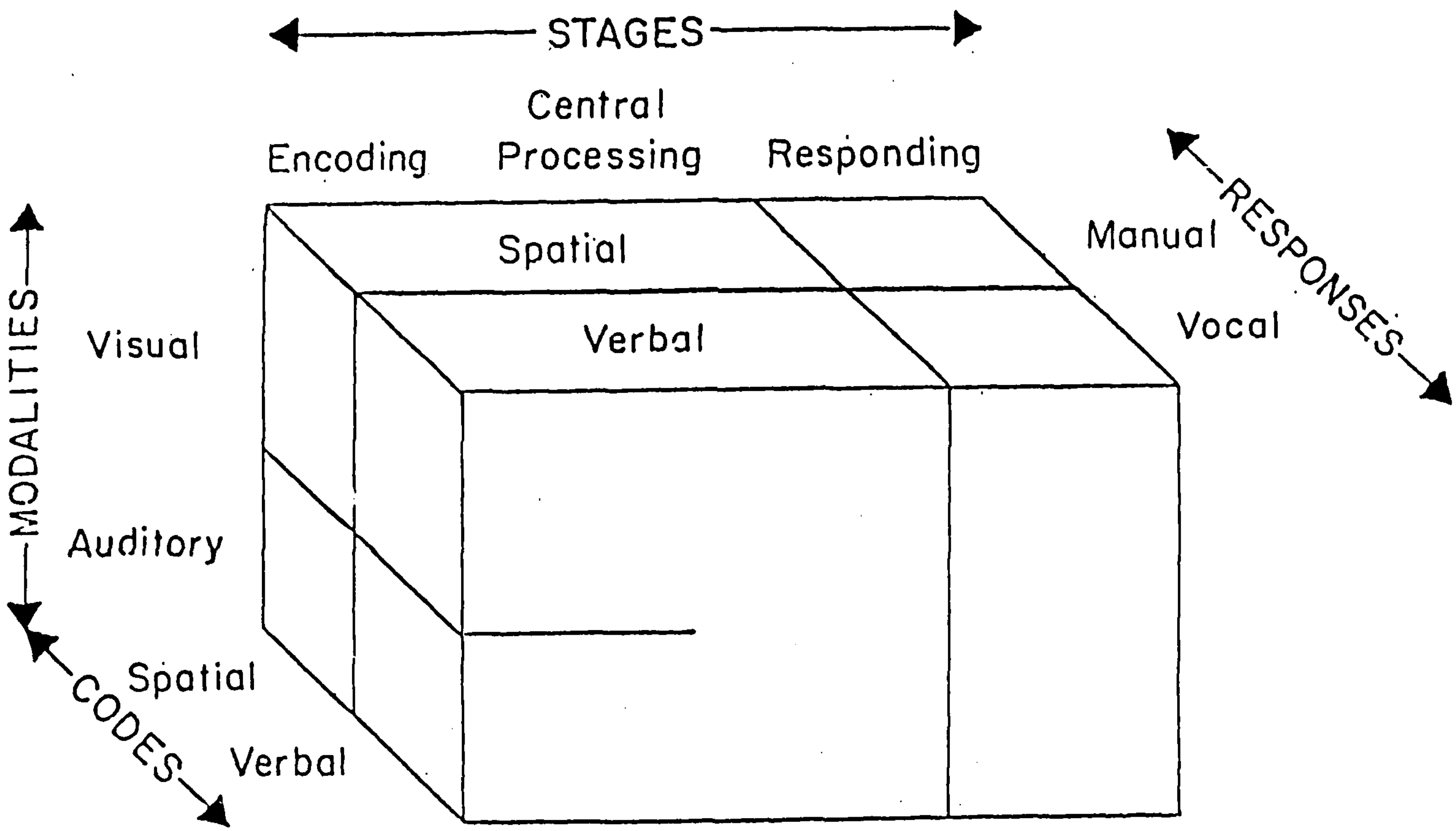


Figure 1. Multiple Resources Model (from Sandry and Wickens, 1982)

spatial code, then a manual response induces higher S-C-R compatibility than if a vocal response had been required. Likewise, a vocal response is more compatible with verbal information than a manual response. As shown by the fact that "modality" is on the vertical axis, the preceding statements hold true whether the information was encoded visually or auditorially. In designing a control system, one should strive to obtain the maximum amount of S-C-R compatibility possible to achieve greater efficiency and accuracy.

Regarding dual task performance, the model suggests and predicts relative performance levels based on interference and competition between and within the various resource pools. The primary implication is that the more two tasks overlap in the pools they need to draw resources from, the more the interference that will occur. The more that the two tasks differ in what pools they must draw from, the more compatible they will be. Thus, if one task requires visual encoding of spatial information, the required response should be manual. And in this case the other task should be designed to require auditory encoding of verbal information, followed by a vocal response. Two goals have been accomplished with such a design. First, the encoding and central processing stages of each task have been made most compatible with the respective respor stages. Second, the

two tasks have been designed to draw from completely different sets of resource pools, which has in turn minimized the predicted task interference. In terms of multiple resource theory, this design has set the processes up so that the various capacities can be devoted to one particular process; they need not be distributed across multiple processes.

The multiple resource theory and its structuring as shown in the S-C-R compatibility model provide the designer of a dual task system with some very potent and reliable guidelines for building a highly compatible human-machine interface. As discussed in Sandry and Wickens (1982), Patsfall (1981), Wickens, Sandry, and Vidulich (1983), and Wickens, Vidulich, Sandry, and Schiflett (1981) these concepts have been applied directly to the application of the pilot-cockpit interface. The study described in this paper also uses these theories as bases for performance prediction.

Norman (1982) discusses the importance of a "direct relationship" between the conceptual model and the operator's mental model of a system stating that this is an essential aspect of a good person-machine interface. A conceptual model is a description of the system provided to the user in an attempt to clearly and accurately represent

the structure and dynamics of that system. A mental model is the description of the system which the user has developed in his or her own mind upon which most decisions regarding operation of the system are made. The user's mental model is developed through training and through experience in operating the system. Therefore, assuming that the conceptual model is accurate, a goal of the training program is to convey the conceptual model in such a way that it can easily be internalized by the user resulting in a direct relationship between the two. Subsequently, during actual operation of the system, any information provided to the user by the system ought to be structured in a manner that will both correspond to and further develop (correctly) the user's mental model. Two important questions are raised. How should the conceptual model be conveyed? In terms of compatibility with the user's mental model, how should the machine display information to that user?

Thomas (1983) provides a brief description of spatial versus serial memory systems which helps serve as background in determining the optimal answers to the above questions. Thomas points out that in memory tests, pictures (spatial information) were more consistently retained than their corresponding labels (serial, or verbal, information). Three possible explanations of the serial/spatial

differences in memory are given: 1) processing level model, in which spatial and serial processing goes through the same steps, but the individual spatial steps are quicker and more efficient; 2) sensory semantic model in which spatial processing takes fewer transformations than serial processing, thus is more efficient and requires less mental workload; 3) dual encoding model which states that spatial encoding generates both spatial and serial codes but serial encoding only generates serial codes; thus spatial encoding induces better memory characteristics. Whichever model is accepted, the spatial presentation of information should incur quicker responses from memory.

Hollan (1984) contends that pictorial displays are more compatible with the subject's mental model of a system. With the direct mapping of the picture onto the mental model, there does not need to be the transformation from words to this spatial model. Hollan describes the STEAMER project which used this fundamental principle in designing an object-based training system for process control. STEAMER helps the subject to develop mental models by providing graphic displays of the system. Hollan argues that resulting representation of the system developed by the subject is more like an expert's representation. Therefore, the subjects should be able to interact more efficiently with the system as a result of the consistency between their

spatial mental models, the conceptual model, and the physical system itself.

Another application of the mental model concept is discussed by Eberts and Schneider (1980). They investigated using computer generated spatial displays to help make human operation of a second order control system an automatic process instead of a controlled process. In a controlled process, which is relatively slow, the subject consciously allocates resources to the task at hand. In an automatic process, which is much faster, the subject does not exercise conscious control over the process. An example is tracking a runway on final approach. If this were a controlled process, the pilot would not be able to react quickly enough because each input would have to be carefully planned, executed, and analyzed upon completion for its success. Before these steps were completed, a new error correction input would be required. Soon the pilot would have the airplane on a divergent flight path. However, since the task has been internalized by the pilot, the steps are accomplished automatically, or subconsciously. Automaticity implies this internalization of the task.

Spatial displays can minimize the amount of effort a subject needs to put into thinking about a spatially-oriented manual response; even to the point where

the thinking is subconscious, the response is automatic, and the task is internalized. To cultivate this type of thinking, a sound mental model must have been developed by the subject. Eberts (1984) expounds further on using spatial displays to enhance subjects' mental models of second-order systems and thus to enhance their control performance and problem solving abilities.

In a complicated system, an efficient man-machine interface requires that the human operator develop a sound, accurate mental model of the system. When this requirement is met, shorter response times can be elicited from the operator because he has a clearer understanding of where the problem lies in relation to the rest of the system which helps to limit the number of optional solution approaches. It also requires that the information transfer be compatible with the operator's model. Fulfillment of this requirement allows the operator to recognize and categorize the incoming information more readily than if it must be transformed to fit into his model. For example, consider the system to be an airplane and the operator its pilot. The pilot must have an accurate spatial model of the aircraft to comprehend the on-board location of an in-flight problem. She must also know, spatially, how that impaired subsystem is related to the other subsystems to predict any possible interactions. As Reising and Kopala (1982) state, the pilot needs to

control her aircraft and its individual systems as opposed to controlling the computer. Thus the computer must be transparent; the pilot must feel an interaction with the plane, not the computer. Also, spatial displays could be used when the on-board diagnostic computer wants to notify the pilot of a subsystem problem.

One possible way to display the faulty subsystem and its relationship to other systems which might be affected is through a hierarchical sequence of displays. This has the advantage of decreasing the amount of information presented to the pilot in comparison to a display which gives all the information at once.

Information theory (see Kantowitz and Sorkin, 1983 and Wickens, 1984 for more detailed discussions) provides a method for quantifying the amount of information transmitted from a source to a receiver. In simplest terms, the theory states that if one has N equiprobable alternatives to choose from, then the amount of information contained is

$$H = \log(2)\ N\ .$$

Thus if one has eight alternatives, this represents three "bits" of information. Double the amount of alternatives to sixteen, and you have four bits of information. If the receiver is told the correct answer out of sixteen equiprobable alternative answers, then she has received four bits of information. Other ways to change the amount of

information is to change the probability distribution of the alternatives, and to provide context (thus decreasing the amount of information at each level).

As a specific example, let us imagine a military fighter aircraft flying through hostile enemy airspace. The pilot is undoubtedly experiencing a good deal of stress. This aircraft is a new model, and correspondingly it still has a number of bugs which haven't yet been completely worked out. However, it has been found that when a problem does occur, it seems to always be among the same set of eighteen problems. Also, these eighteen problems seem to occur equally often across the group of aircraft. In their training, the pilots have been warned that it is likely that they might encounter one or more of the problems during their mission, and that any one problem is just as likely as any of the other seventeen. To return to the example mission, a warning tone has just notified the pilot that he has a problem. It could be one of eighteen possibilities, but which one? The on-board computer can tell him which one in a variety of ways. One way would be to tell the pilot right out what the problem is. In this case, the pilot has received 4.170 bits of information because log(2) 18 = 4.170. As his attention is already occupied with the problem of flying through hostile airspace, this is a lot of information to load into his short term memory which is

limited to about seven chunks (Miller, 1956).

The computer can be designed to decrease this overload. To do this, the computer provides the pilot with a series of displays which zero in on the information following the hierarchical path from the apex ("emergency") to the specific problem ("left engine fire"). The set of engines, or propulsion system, is one of three systems in which the problem might occur. The computer tells the pilot, "propulsion," transmitting log(2) 3 = 1.585 bits of information. The left engine is one of three engine combinations in the propulsion system which might have a problem, so when the computer tells the pilot, "left engine," it again transmits log(2) 3 = 1.585 bits of information. Finally, a fire is one of two problems that might occur in the left engine of the propulsion system. By telling the pilot, "fire," log(2) 2 = 1.00 bit of information is sent. As can be seen, each time the computer gave the pilot some information, the bits of information or the uncertainty was reduced and, therefore, less demand was placed on the pilot's processing. This which has two immediate benefits. First, he can process the emergency information more quickly and second, he uses up fewer of the resources that should be allocated to that other important task: flying the plane.

The impact of incorporating information theory into display design is enhanced by a theory base regarding organization of the human memory. Memory is often characterized as being organized hierarchically (e.g. Mandler, 1968). If a set of elements (words, actions, responses, etc.) are to be committed to long term memory, they should be associated and categorized hierarchically to fit in the mental organization. If we follow the path from an element in the "bottom level" of the hierarchy up to the apex, each element encountered along the way can be considered as a level of context for the elements below. Thus subsequent recall of an element at the bottom level will be facilitated if the elements along the downward path from the apex (ordered context) are presented sequentially.

The hierarchical model has been the topic of much research, both in studies relating to simple word recall and in more complex human-computer interactions. Most of the studies have supported the model, though some theoreticians have suggested alternative schemes. Bower, Clark, Lesgold, and Winzenz (1969) demonstrate conclusively in a series of five word recall experiments that words presented in a meaningful hierarchy were much more readily recalled than when presented in a random hierarchical structure. Summarizing their findings, the authors state that if a

subject finds a simple relationship between the words in a list, then that relationship can be used to help retrieve the words from memory resulting in better performance of the memory task. The relationships used by the subjects in these experiments were associative hierarchies.

Broadbent, Cooper, and Broadbent (1978) test the hierarchical model against a non-organized scheme in word recall. In this experiment they derive results similar to those of Bower et al. (1968). However, a further investigation in which they compare a hierarchical scheme to a "matrix" scheme brings them to the conclusion that the matrix scheme may sometimes be as good as the hierarchy.

While these two studies supported the hierarchical model of mental organization through word recall tests, there have also been a number of studies which apply this model to the domain of human-computer interaction problems. For example, Liebelt, McDonald, Stone, and Karat (1982) and Miller (1981) apply the model to computer menu structures. Liebelt et al. confirm the advantages of a pure hierarchical menu structure, while Miller hypothesizes on the optimal size, "depth", and "breadth" of the hierarchy. These two studies pertain to the general field of human-computer interaction, and therefore specific applications should also follow the guidelines produced. In fact, in the conclusion

of Miller's (1981) article, he does suggest that his results could be applied to specific situations such as the military cockpit.

The hierarchical theory has been shown in many cases to apply to very specific interfaces. Dray, Ogden, and Vestewig (1981) analyze the application of hierarchical menus to the Stand-Off Target Acquisition System (SOTAS) which is a computer-controlled weapon system intended for use aboard Army attack helicopters. This study demonstrated the advantages of learning characteristics provided by the menu structure. Henneman and Rouse (1983) study the depth-breadth trade off in menu display of a telephone network process control system. These studies all have incorporated an obvious hierarchical organization as a way of decreasing the response time and increasing the response accuracy of the subjects involved. As stated earlier, the concept of context is closely related to that of hierarchies. At each level of a hierarchy, context is given which directs the operator to the proper area of the next lower level in the hierarchy.

The Simpson and Williams (1980) study discussed previously addressed the context question. As they found, providing more context improved the pilot's performance and possibly even lowered his mental workload. After given the

first context word of the warning message, the pilot had fewer alternatives for what the following word might be; the first word had directed him to a more specific location of the hierarchy. Again, Hakkinien and Williges (1982) take things one step further and show that an alerting tone preceding the warning messages acts as one more level of hierarchical context.

Rouse (1984) suggests that in familiar but infrequent situations (such as cockpit emergencies) information should be presented in a "disaggregated" format. This allows the operator to match the pieces of information to his own mental model of the system and display/response relationship. Since this mental model is referred to infrequently and under high stress, the information matching needs to be done in a series of steps instead of in one display. The series should then follow a hierarchical format to be most compatible with the pilot's organization of the response information.

Presenting the information in such a hierarchical format may indeed be a valuable alternative to presenting it all in one display. Verbal information is serial by nature; it inherently reduces the uncertainty as the information is presented. Perhaps this is why verbal information has been so good in the past. To compare spatial information with

verbal information, the spatial information should be presented serially also. Naturally there is a trade-off; if the message consisted of too many levels of context then the plane might explode before the pilot gets the whole message. On the other hand, if the pilot has to decode an overcrowded picture, the plane might explode before he finishes, or dive into the ground because he is concentrating so hard on processing all the information.

Human information processing has received much theoretical attention which has resulted in a variety of models representing different aspects of human performance. While no single model can describe every aspect of information processing, a good combination of ideas from the different models can help in finding the optimal solution for a specific application. The pilot-cockpit interface is one which involves multiple simultaneous tasks, high mental workload, and quick, accurate decisions and responses. A set of guidelines to help meet these demands can be derived from the theoretical premises of multiple resources, S-C-R compatibility, mental models, information theory, and hierarchical mental organization.

The Problem

The question of how the on-board computer ought to display information to the pilot during emergency situations

is presently an important topic since technological advances have introduced two distinct alternative methods. These are the CRT or flat panel displays, and digital speech generation. The question has been approached from both empirical and theoretical viewpoints but as yet an optimal display method has not been agreed upon. Flying an aircraft is a task in which the pilot encodes and processes spatial information through the visual modality, and responds manually. Current multiple resource and S-C-R theory suggests then that secondary tasks (such as responding to emergencies) should utilize the diametrically opposite resource pools. This would include encoding and processing verbal information through the auditory modality, and responding vocally. Curiously, though, as described in an earlier section of this paper, speech I/O has not consistently outperformed visual/manual I/O in secondary task performance even when the primary task was visual/manual.

What is it about pictorial displays that allows them to elicit nearly equal performance as speech displays when, from one theory, they should not? In the previous research, the pictures were not fully optimized from the theoretical viewpoints discussed earlier. Consideration of the other two concepts discussed, mental models and hierarchical structuring, may reveal some valuable insights. The modern

aircraft is an extremely intricate system. When an emergency occurs, the pilot needs to be able to consult a spatial model as this may allow much quicker mental scanning of the system than does a verbally (serially) constructed model. Pictorial displays may not only bolster the development of an accurate spatial mental model, they may also present information which is more compatible (thus more efficiently processed) with the pilot's mental model.

Another way to help the pilot mentally scan the system quickly is to "zoom in" on the fault location and description. This approach has been shown to improve performance in studies attempting to optimize speech displays; hierarchical context appeared to decrease the pilot's mental workload. Studies involving pictorial displays have not utilized this concept extensively. Instead, large amounts of information have been placed on one display which not only clutters it but also requires finer detail. A series of quick glances at the screen while it is zooming in on the problem with larger, less detailed pictures should have the same effect as hierarchical context provided vocally.

Finally, spatial displays may be better than vocal displays in another aspect. The concept of stimulus-response compatibility was demonstrated by Fitts

and Seeger (1953): if the proper response to a particular condition is on the left side of the control panel, then the display should reflect this by directing the subject's attention to the left side of the display. This is one concept that has not been sufficiently implemented in studies comparing speech to pictorial display.

Perhaps the S-R compatibility theory conflicts with the multiple resource and S-C-R compatibility theories discussed above. Assume a primary task in which the encoding utilizes visual and spatial resource pools, spatial pools for central processing, and manual responses. If a secondary task is added which utlizes the same resource pools, then there is a good chance that these pools will become overloaded. Now assume that the secondary task, while still including manual responses, utlizes auditory and verbal resource pools for the encoding stage and verbal resources at the central processing stage. This setup is good because it spreads the two tasks over different pools in the first stages of processing, but then the crossover to manual responses in the secondary task can cause interference.

It is easier to incorporate a direct mapping between pictorial displays and required responses than between speech displays and the responses. The problem is, does the advantage of spreading the tasks over the resource pools

outweigh the advantage of high S-R compatibility available in spatial/pictorial displays?

As suggested at the beginning of this paper, for a number of reasons, generated speech displays have been attracting more attention than pictorial displays. Most of the reasons for using speech displays are theoretically sound, but perhaps not theoretically complete. It is essential that we make sure to utilize all the possible advantages of pictorial displays when comparing them to speech displays, otherwise the comparison is invalid.

The purpose of the proposed study is to compare the advantages of pictorial emergency displays to generated speech displays. In particular, both types of displays will incorporate hierarchical structuring and the pictorial displays will be designed to be compatible with the structure of the response panel. It is expected that because of the spatial relationships inherent in the pictures, subjects receiving pictorial displays will develop stronger and more useful mental models of the system and the stimulus - response interface than subjects receiving speech displays. Even though the subjects receiving pictorial displays must draw resources from the same encoding and central processing pools used for the flying task, the processing-response compatibility and better model will

outweigh the advantage of the speech subjects (who need not draw from the same encoding and processing pools).

THE EXPERIMENTS

Three experiments were conducted to test the advantages of spatial characteristics in pictorial displays. In all three experiments, the effects of display presentation modality (speech versus pictorial) on pilot performance was studied. Performance was measured in terms of emergency response time and accuracy as well as flying performance. The other variable of interest in all three experiments was task type; whether or not the spatial advantages in pictorial displays are apparent in dual task as well as single task situations. In each of the three experiments, a different third parameter was varied to study its main effects and its interactions with modality and task type. The primary factor of interest is the display modality. As was stated previously, a main concern in all experiments is the possibility that the direct mapping from pictorial display to response is as helpful as utilizing different processing modalities as speech does.

However, these variables considered alone may not show all of the advantages associated with either of the display methods. Interactions with other variables can also show

advantages; for example responses to one display method might be more easily learned than to the other. Thus the primary purpose of including three different experiments is to allow the analysis of potential interactions which may impact a decision on emergency display application. Table 1 shows, for each experiment, what the third variable is and why it is included in the study.

Table 1. The Third Variable and its Purpose in Each Experiment

Experiment	Variable	Purpose
One	Practice	To determine if pictorial displays might help subjects learn the display-response relationship more quickly than speech displays.
Two	Message Rate	To determine the effects of varying the rate at which messages are presented; to find if there are any interactions with display type that might need consideration in the applications of the displays.
Three	Labels	To determine if the pictorial displays helped subjects build less dependency on the response labels; if their internalization of the S-R relationship is more helpful than when speech displays are used.

The experiments, all three of which each subject participated in, followed the same basic method. Therefore a detailed description of the method will be presented in the "Experiment One" section, with any respective

differences noted in the sections describing experiments two and three.

Experiment One

The primary motivation of Experiment 1 was to examine the effects of practice and its interaction with display type. If, as discussed in the introduction, the pictorial subjects develop internal representations of the S-R compatibility more quickly than speech subjects, an interaction between practice and display type should occur. This might suggest that the direct spatial mapping from stimulus to response might provide advantages which are equally or more important than the distribution of input modalities over processing resources.

Method

To provide a realistic paradigm for gathering data, the experiment used emergency conditions during flight in a fighter cockpit. The tasks consisted of 1) flying a cockpit mockup through hostile territory, and 2) responding to on-board emergencies such as engine fires and hydraulic failures. The main treatment was input modality which considered two modality/code combinations; auditory/verbal and visual/spatial. As stated earlier, the other parameters were practice and task type. The twenty subjects were required to perform two single task missions and one dual

task mission. This procedure was repeated to examine the effects of practice.

For simulation of the fighter cockpit, a fixed-base F-16 mockup was used. The primary task was a tracking task which simulated the actual mission which the pilot was to fly. For the secondary task, the subjects had to respond to various emergencies which occurred during the missions. These emergencies were critical; failure to respond immediately would have serious consequences in a real aircraft. This dual-task setup allowed for measurements of the effects of each task in a high workload, high stress situation. There may be some controversy as to which task really ought to be considered the "primary" one and which the "secondary" one. It may seem as though the response task ought to be considered as the primary task since that task is the one upon which the treatments are varied; or as Navon and Gopher (1979) put it, the difficulty of the response task is varied. When the subjects were trained, they were told that immediate response to the emergencies was of utmost importance, in both the single task and the dual task runs. Thus one might infer that the response performance ought to be held constant: maximum speed and accuracy at all times. However, in the theory/reality tradeoff of this experiment, it was necessary to consider the priority rules which are part of every pilot's training

in an emergency. As outlined in the F-16 Operating Manual (1979), the top priority is to "Maintain Aircraft Control". The second and third priorities are to "Analyze the Situation and Take Proper Action", and to "Land as the Situation Dictates". This suggests that the most important task is to keep flying the plane and as soon as possible, attend to the emergency. Even in a hostile environment for example, the pilot should first control the aircraft, evade an enemy missile, and then tend to the emergency. Or in other words, keep the performance of the flying task constant while attending to the emergency; make the flying task the "primary" task. This was the reasoning followed for selection of task designations for this experiment. As stated previously, it was expected that performance of the primary task would, however, degrade significantly with addition of the secondary task.

In planning the experiments, it was foreseen that each subject would participate for three to four nearly continuous hours. It was felt that for this length of time a conventional tracking task would be tiresome and non-motivating for the subjects. A viable alternative was to use a home arcade video game which would be intrinsically motivating for the subject throughout the full test period. This approach has been used before, for example see Williamson and Curry (1984).

The primary task consisted of "flying" an aircraft through hostile territory; avoiding enemy surface-to-air missiles, stationary ground obstacles, enemy interceptor aircraft and its gunfire. Meanwhile, the subject had at his disposal an unlimited supply of forward firing missiles and gravity bombs with which he could gain points by destroying enemy targets. This realistic attack mission was provided by the commercially available "Cosmic Avenger" video game cartridge made by ColecoVision. The game mission actually includes three different types of territory through which the pilot must fly.

In the first part of the mission, the pilot finds himself flying over a fortified city which is heavily guarded with surface-to-air missiles (SAMs) and anti-aircraft flack bombs. Two types of SAMs, pursuit and non-pursuit, are encountered by the pilot. When the pilot flies over a pursuit type SAM, the missile takes off at a 45 degree angle until it reaches the altitude at which the pilot is flying. When it reaches this altitude, the SAM levels out, accelerates, and approaches the pilot from behind. These SAMs are "smart"; if the pilot inputs an altitude change, the SAM will respond by correcting its altitude to that of the pilot. This correction, however, follows a short time lag. Thus a possible evasion maneuver

for the pilot is to wait until the missile has nearly caught up to him, then "duck" under or over the missile, pull back on the throttle, and let the missile fly by. (The pilot can then score a hit on the missile from behind with his own missiles.) The non-pursuit missiles are not "smart"; they simply launch vertically as the pilot approaches. The pilot must maneuver to avoid these missiles or shoot then down with his on-board missiles.

The other surface-to-air obstacle encountered is the "flack-bomb". This is a projectile which is launched vertically and at some altitude explodes, dispersing "flack" or shrapnel over a wide area. If the pilot flies though the flack, his plane is destroyed. The explosion altitude of the flack bombs is not known by the pilot beforehand. Thus when approaching the rising flack bomb the pilot must take a risk in deciding whether to fly above or below the bomb. He also has the opportunity to shoot down the flack bomb before it explodes. These enemy projectiles are not too difficult to deal with individually, but the pilot is rarely in a one-on-one situation. Usually he has to contend with many of the missiles simultaneously, making the task much more difficult. And, to add to the difficulty, a persistent force of enemy interceptor aircraft does its best to deprive the pilot of his airplane, and his life.

These interceptors attack the pilot one at a time. They fly at high velocity and their flight paths are highly irregular and unpredictable, thereby making it extremely difficult for the pilot to keep from running into them (let alone to shoot them down). To make things worse the interceptors are armed with missiles, the erratic firing of which often catches the pilot off guard. The pilot is provided with a "radar" display at the top of the screen which allows him to locate these interceptors one screen width ahead or behind the displayed screen.

In the second part of the mission, the pilot leaves the cityscape and flies out over barren "plains" which are crawling with tank-like vehicles. The tanks are, of course, equipped with anti-aircraft artillery so while the pilot is trying to "kill" the tanks, he must avoid the constant barrage of artillery fire. To make matters more interesting for the pilot, the interceptor aircraft encountered in part one have no qualms about extending their effectiveness into part two of the mission.

In the third mission section, the pilot enters a scenario resembling underwater caverns. The roofs and floors of the caverns are irregular, and at times the passage between these is quite narrow. The pilot must avoid or shoot down many passive mines as well as stationary

submarines which shoot torpedoes at him. He must also contend with missiles similar to the "smart" SAMs described in part one, though they approach him head-on in this stage. When (if) the pilot emerges from the caverns, he finds himself once again in the "cityscape" environment, but this time the ground level has been raised which gives him less maneuvering space thus increasing the difficulty of the task. Each subsequent time that the pilot successfully negotiates the three mission parts, the difficulty level is increased in the same manner.

While a major goal of the mission is simply "staying alive", the other major goal consists of destroying as many of the enemy targets as possible. All flying objects are considered targets as are all ground-based facilities such as SAMs which have not yet been launched. As mentioned previously the pilot can destroy these targets using either gravity bombs or forward-shooting missiles. Not only did destroying targets improve the pilot's chances of survival, but he was awarded points for his "hits". The score display on the screen provided the subject with more motivation to perform well, i.e. to better his score from the last run.

With this tracking game, the subject was loaded with a task not unlike those encountered by pilots in actual attack missions. Since this task demanded a good deal of

processing resources, the subject had to devote much of his attention to it for successful performance. Unfortunately for the pilot, not only did he have to face relentless conditions imposed by the enemy, but he also had to contend with his own aircraft which turned out to be quite unreliable. There were frequent emergencies regarding his on-board systems to which he had to react in a timely manner to stay alive. Thus the pilot was forced to direct some of his attention, or processing resources, away from the flying task toward the emergencies.

While flying the simulator, the subject often ran into problems with his own aircraft such as engine fires, electrical power-outs, and hydraulic pump failures. As these conditions imposed serious threats to his survival, it was imperative that he respond as quickly as possible by pushing an appropriate button such as the fire extinguisher control. Perceiving, processing, and responding to the emergency information which the on-board computer provided him with, then, constituted the secondary task. It was this secondary information which received the various treatments to determine how the pilot's performance would be affected. As stated earlier, the main treatment was input modality and other parameters were practice and task type.

During the training session, the pilots were told that their plane was equipped with an on-board computer which was very good at diagnostics. When a system had a problem the computer would diagnose it and present the diagnosis to the pilot so that he could initiate the remedy for the problem. The subjects were also told that while the computer was very good at diagnostics, the aircraft designers had decided that the computer should not automatically initiate the fix; the pilot was to be the mission executive and there might be times when he would not want an immediate fix. For example, if the pilot was flying a tight maneuver to evade an approaching missile and an engine caught fire, he might need one more second of thrust from that engine to dodge the missile before shutting the engine down and blowing the fire extinguisher. If the computer had initiated the shutdown immediately, the pilot might not have enough thrust for effective evasion and would be in worse shape than if the engine had been allowed to burn one second longer. Therefore, the computer would only tell the pilot about the problem and leave it up to him to take appropriate action.

In describing the emergency to the pilot, the computer presented a hierarchical sequence of four displays. The format proceeded from general area to specific problem, thus "zeroing in" on the exact problem. In all cases, when an emergency occurred the computer notified the subject of the

impending message by issuing a .5 second beep. The first stage of the message was the "warning" stage - this notified the pilot that the incoming information concerned an emergency status. The second stage was the "main system" stage - here the emergency was narrowed down to one of three systems: the hydraulic, the electrical, or the propulsion system. Following this was the "subsystem" stage - this stage narrowed the problem further to the left, right, or both subsystems. Finally, the "malfunction" stage narrowed the emergency down to one of two possible malfunctions in the faulty subsystem of the defective main system. Thus instead of having to discern between eighteen possible emergencies, the subject had to discern at most between three alternatives at each level. Figure 2 displays the hierarchical relationship of the emergencies, subsystems, and main systems.

Two types of displays were used to present the emergency information to the subjects: 1) digitized speech, and 2) pictorial display. The term "modality" will be used in this paper to indicate the display type parameter.

The digitized speech output came from a speaker positioned on the left side of the cockpit mockup. Following the message notification beep, the word "Warning" was issued from the speaker. After this, three more one- or

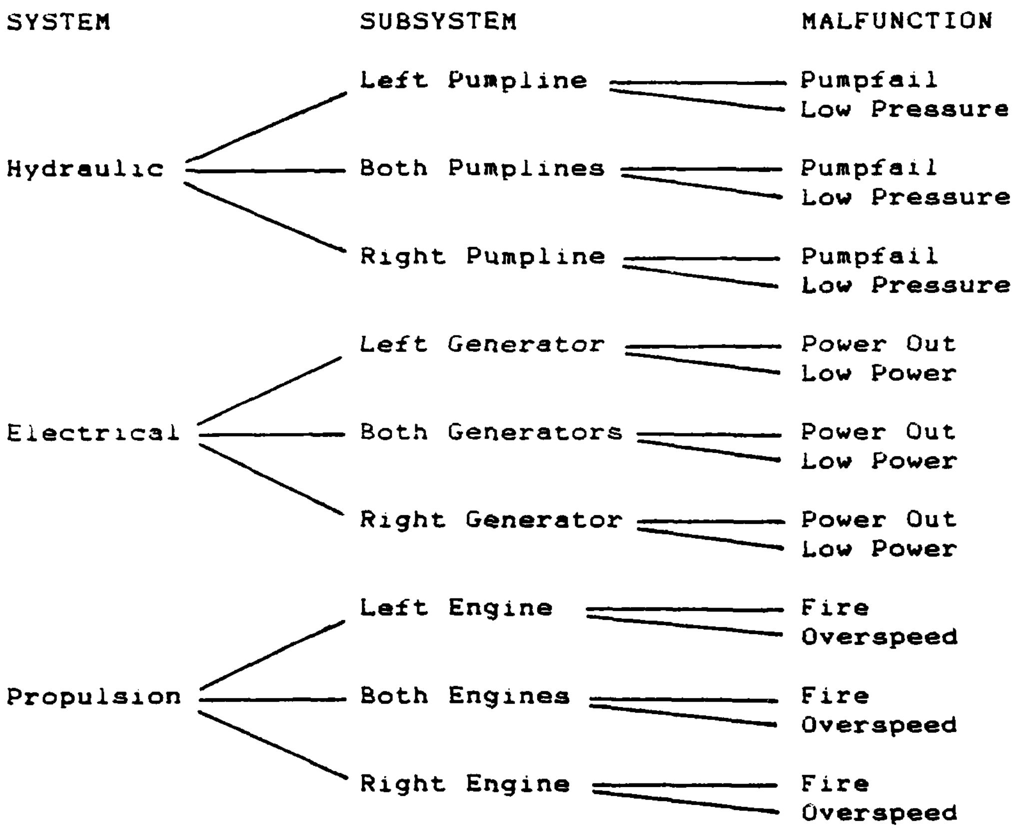


Figure 2. Hierarchical Structure of Emergency Warning Messages

two-word phrases as shown in Figure 2 were heard.

The pictorial displays (see samples in Appendix A) were back-projected onto a screen below the video game display. They followed the same sequences shown in Figure 2, i.e. instead of hearing four phrases the subject saw a series of four pictures one at a time, paced by the projector.

The response keyboard consisted of eighteen keys; each dedicated to one of the eighteen emergencies. The arrangement of the keys corresponded to the grouping evident in the hierarchical format, as Figure 3 depicts, and also corresponds to the spatial location and severity of the problem. For example, response buttons dealing with emergencies in the Electrical System were grouped together, and response buttons dedicated to "left" subsystems were located on the left side of the keyboard. For the speech subjects, the buttons were labelled verbally (as in Figure 3); for the pictorial subjects the words were replaced with pictures corresponding to those seen on the CRT display. To acknowledge the subject's response input, the computer issued a .2 second blip when the subject hit a button. This blip was different from the message notification beep - higher frequency and shorter duration - so that the subject would not confuse the two, thinking that a new emergency had come up when he hit the response button. The subject was

		LEFT	BOTH	RIGHT
HYDRAULIC	PUMPFAIL			
	LOW PRESSURE			
ELECTRICAL	POWER OUT			
	LOW POWER			
PROPULSION	FIRE			
	OVERSPEED			

Figure 3. Response Panel and Label Configuration

limited to one button push for each emergency; the computer ignored subsequent pushes and no blip occurred after these. Thus if the subject realized he had made a mistake he could not correct it by pushing the proper button.

The primary operational equipment used in this study consisted of the F-16 fighter mockup cockpit illustrated in Figure 4. The video game was displayed on a CRT located in a typical Heads Up Display (HUD) position of the cockpit. The subjects had two manual controls for the game. In the right hand was the altitude control and in the left hand was the throttle. The weapon firing buttons (missile and bomb) were both located on the altitude control stick.

For the pictorial displays, slides were back-projected onto a ground-glass screen located below the HUD in the center of the forward cockpit panel. The screen simulated a CRT display in an actual cockpit. The individual pictures were originally composed on a Texas Instruments Professional Computer using the graphics statements available in the T.I. Basic language. The images were then photographed on color slide film.

For the speech displays, a speaker was located on the left side of the cockpit facing the subject. This speaker was driven by a VOTAN V5000A digital speech generation

Figure 4. Fighter Cockpit Mockup

system. The individual phrases (corresponding to the individual slides in the pictorial displays) had been pre-digitized and stored in the system memory.

The test operator's control console was located behind the cockpit mockup as shown in Figure 5. The console consisted of the control computer interface and a parallel CRT displaying the video game which the subject was "flying." The control interface included a CRT display and a keyboard for the operator to enter various test control commands, parameter levels, and inputs to initiate the emergencies. The control CRT displayed such information as the current test matrix number, current emergency, proper and actual subject responses, and subject error flags.

Twenty male subjects participated in the study. All subjects were employees of Wright-Patterson AFB, OH, and all either had at least a bachelor degree in science or engineering, or were working toward one. The ages ranged from 19 to 42, with a mean age of 25.3 years. None of the subjects were trained military pilots.

In the beginning of the experimental session, the subject was given a standardized briefing describing the purpose of the experiment and a general description of the tasks that he would be expected to perform. The scripts for

Figure 5. Test Operator's Control Console

the initial briefings given to pictorial and to speech subjects are provided in Appendix B. Following the initial briefing, the subject was given twenty-five minutes to become familiar with the video game. At the end of this period, he was scored for one cycle (five ships) of the game.

Following the single task game run, the subject was given a detailed briefing describing the emergencies that could occur. In this briefing (Appendix B) all the slides were demonstrated on the screen, or if he was in the speech subject all the words were spoken through the speaker one at a time. The subject was also informed about the hierarchical message format and its purpose of zeroing in on the problem. During this time he was familiarized with the response panel and shown which buttons corresponded to the various emergencies. After this briefing and demonstration, the subject was administered a single task (emergency) test during which data was gathered. In this test he was given the eighteen emergencies in a random order, and encouraged to respond as quickly as humanly possible.

With the two single task runs completed, the subject was ready for the first dual task run. In this run, he was required to respond to the emergencies as quickly as possible while playing the video game. However, he was also

told that he should not let up on the video game during an emergency; i.e. to protect his performance of the primary task. The mission was completed when the subject had responded to all eighteen emergencies, again presented once each in a re-randomized order. Most subjects required more than one game to complete the mission; i.e. their first five ships had been killed before receiving all eighteen emergencies. In this case the game was simply reset and the subject was given five new ships.

Following the first dual task mission, the subject was given a second dual task mission, single task (emergencies) mission, and single task (game) mission to test for practice effects. In summary, the order of the runs were as follows:

1. Training -- Video Game
2. Single Task Video Game -- No Practice
3. Training -- Emergency Responses
4. Single Task Emergency Responses -- No Practice
5. Dual Task -- No Practice
6. Dual Task -- Practice
7. Single Task Emergency Responses -- Practice
8. Single Task Video Game -- Practice

Three primary measurements were taken: response time, response accuracy, and game score.

1. Response Time. This was measured, in hundredths of seconds, from onset of the last slide (pictorial) or phrase (speech) of the warning message to the first keystrike on the response panel.
2. Response Accuracy. This was measured by the number of incorrect responses in each mission of eighteen emergency responses. The correct response and the subject's actual response for each emergency was recorded.
3. Game Score. At the end of each five ship game, the final video game score (based on number of enemy targets killed) was recorded. The scores for a mission were totalled and divided by the number of ships used, resulting in a score per ship measure.

Other measures which were recorded included the total number of ships used, the number of ships killed by the enemy during a task (emergency) and those killed between tasks.

The experimental factors were modality, practice, and task type. The experimental design could be classified as a "Nested Factorial", with subjects nested under modality. Practice and task type provided the factorials. The model used for analysis of variance is shown in Appendix E.

Results

The pictorial messages were responded to faster than the speech messages (see Figure 6 and Table 2); this main effect was marginally significant ($F(1,18)=4.138$, $p<.057$). In addition, responses were quicker in the single task setting than in the dual task setting ($F(1,18)=41.969$, $p<.0001$). No significant differences occurred with practice. The modality by practice interaction (see Figure 7) indicated that with practice, the subjects receiving pictorial messages improved in response time more than did the subjects receiving speech messages ($F(1,18)=6.363$, $p<.021$). Running a simple effects test of the modality factor at each of the two levels of practice showed that while mode effects were insignificant with no practice, they were significant with practice, ($F(1,38)=6.3$, $p<.02$).

Only two main effects were found to be significant when measuring response accuracy (see Figure 8 and Table 3). Responses were more accurate in the single task tests than in the dual task tests ($F(1,18)=20.766$, $p<.0002$), and they became more accurate with practice ($F(1,18)=13.722$, $p<.002$). All other main effects and interactions were not significant at the .05 level. An analysis of the types of errors made is in Appendix F. The errors were classified in four groups; 1) left/right subsystem reversal, 2) emergency type (within subsystem) reversal, 3) incorrect system, and 4)

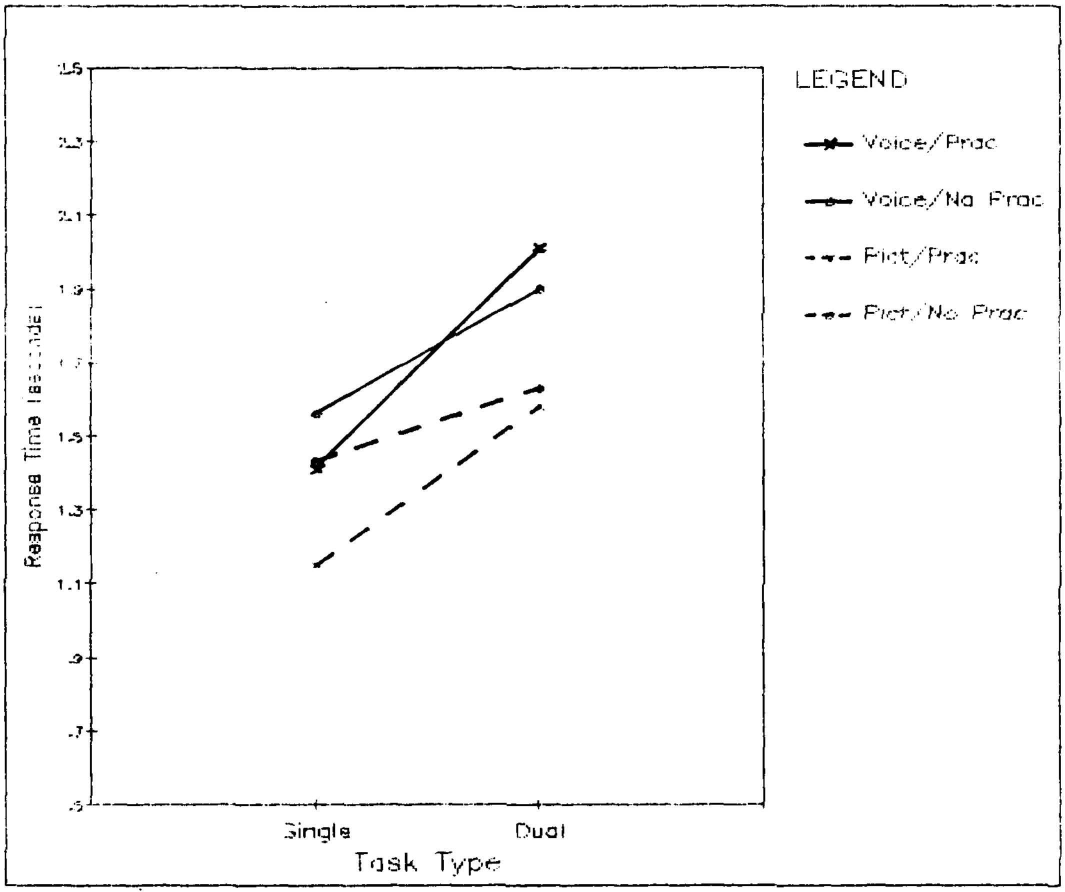


Figure 6. Effects of Modality, Practice, and Task Type on Response Time in Experiment One.

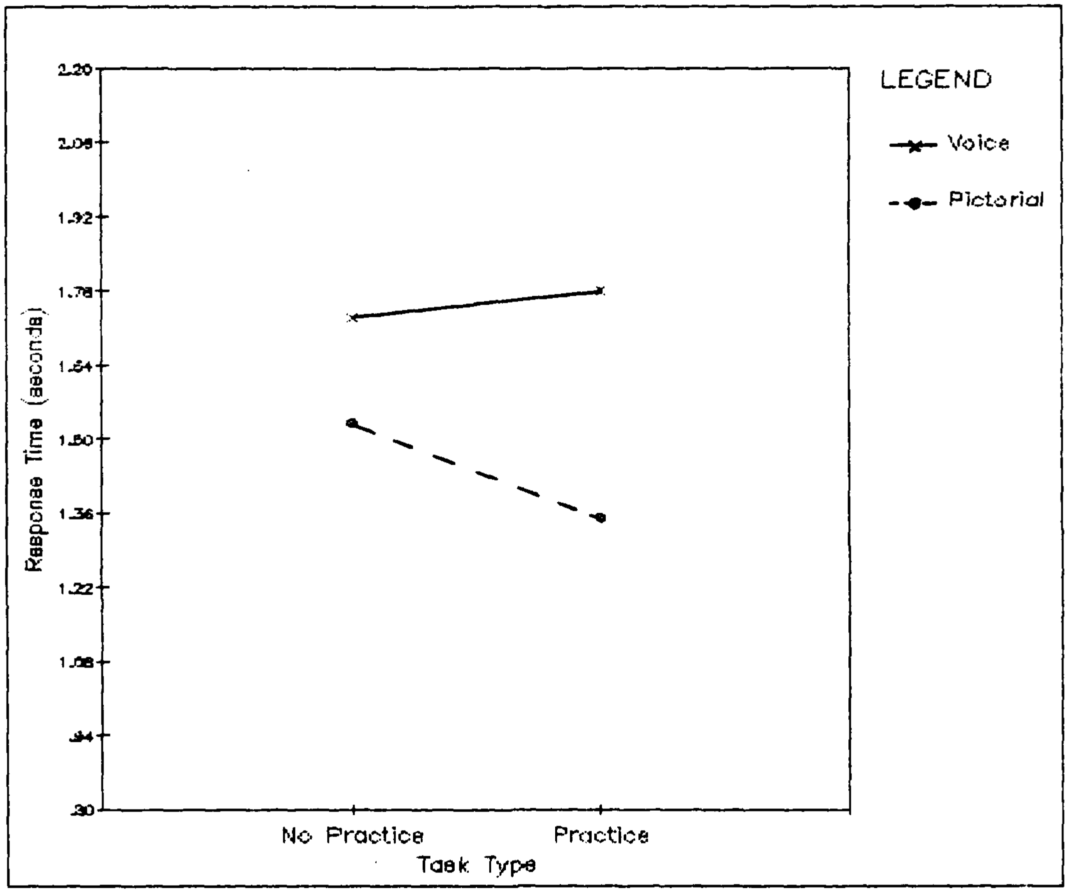


Figure 7. Effect of Interaction between Modality and Practice on Response Time in Experiment One.

Table 2. Significance Tests for Response Time in Experiment One.

SOURCE	DOF	MEAN SQUARE	F	p	
residual	0	-----			
mean	1	205.793	--	--	
modality	1	1.9127	4.138	.057	
error	18	.4627			
task type	1	3.6851	41.969	.000	*
modality X type	1	.26335	2.999	.100	
error	18	.0878			
practice	1	.0714	2.1104	.164	
mod X practice	1	.2153	6.3631	.021	*
error	18	.03383			
type X prac	1	.4789	11.8745	.003	*
modXtypXprac	1	.0357	.8851	.359	
error	18	.04033			

Table 3. Significance Tests for Response Accuracy in Experiment One.

SOURCE	DOF	MEAN SQUARE	F	p	
residual	0				
mean	1	259.200			
modality	1	6.05	1.1678	.294	
error	18	5.181			
task type	1	48.050	20.7659	.0002	*
mod X typ	1	.8000	.3457	.564	
error	18	2.314			
practice	1	16.200	13.7223	.002	*
mod X practice	1	.0500	.0423	.839	
error	18	1.181			
typ X prac	1	.450	.1686	.686	
modXtypXprac	1	5.00	1.873	.188	
error	18	2.669			

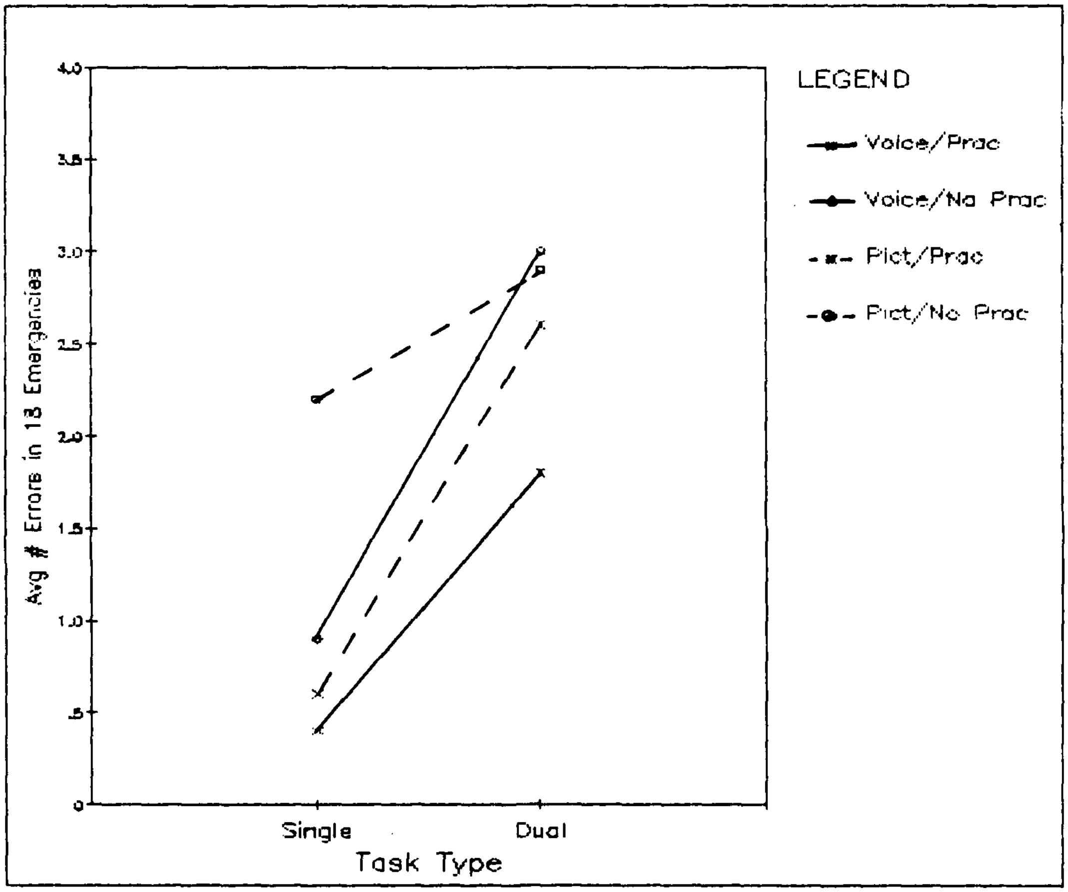


Figure 8. Effects of Modality, Practice, and Task Type on Response Accuracy in Experiment One.

left or right subsystem reversed with both subsystems. The distribution of errors made by the pictorial subjects was significantly different than that made by speech subjects ($X2(3, N=10) = 13.60$, $p<.005$).

For the video game score, the only two significant effects came from Task Type and Practice. Figure 9 (see also Table 4) shows that scores were higher in the single task category than the dual task ($F(1,18)=14.93$, $p<.001$), and they became higher with practice ($F(1,18)=10.64$, $p<.004$. The other main effects and interactions were not significant.

Discussion

The shorter response times associated with the pictorial displays (especially with practice) support the expectation that the stimulus - response compatibility possibly offers more advantages than spreading the two types of input information over two modalities. Based on the multiple resources information processing theory, these shorter times may not be expected. One might think that since the flying task already utilized much of the capacity available from the visual modality and spatial code resource pools, less capacity was available to apply to a visual/spatial secondary task than to an auditory/verbal task. Therefore the responses to the visual/spatial

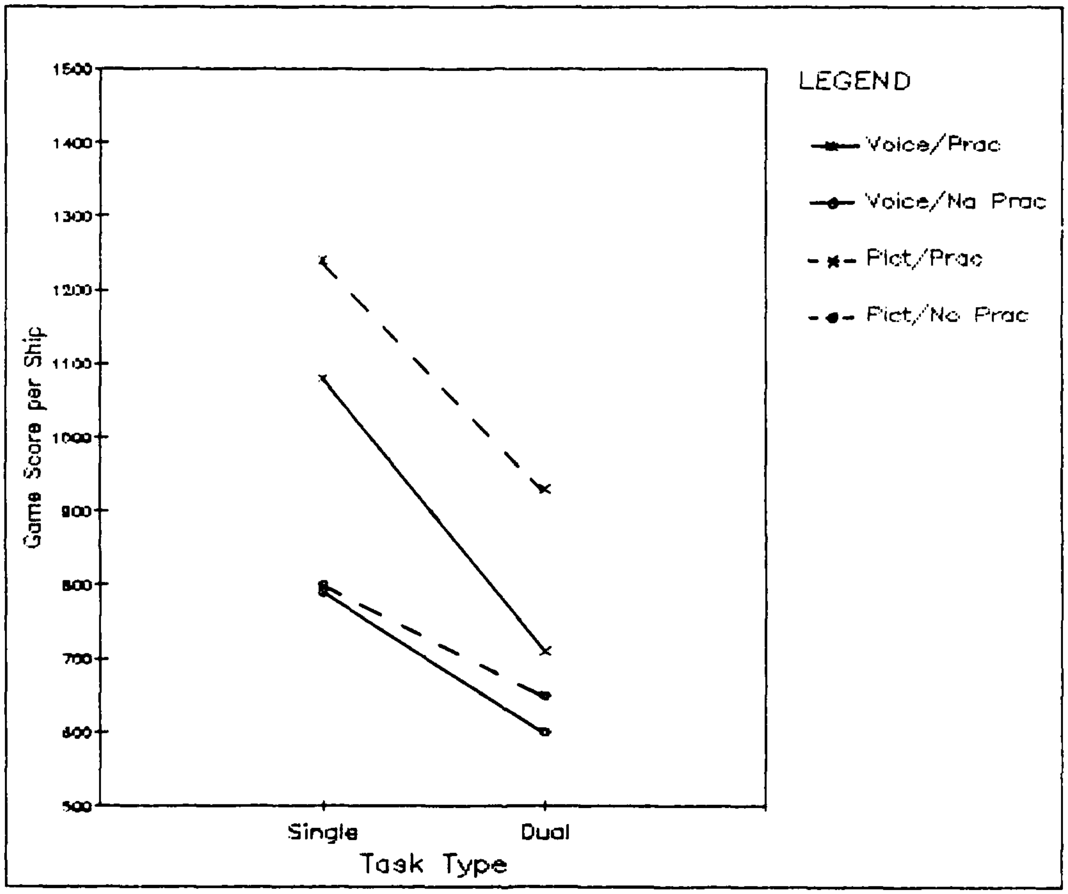


Figure 9. Effects of Modality, Practice, and Task Type on Video Game Score in Experiment One.

Table 4. Significance Tests for Game Score in Experiment One.

SOURCE	DOF	MEAN SQUARE	F	p	
residual	0				
mean	1	1.581E8			
modality	1	244868.	.4728	.500	
error	18	517860.			
task type	1	1.276E6	14.93	.001	*
mod X task type	1	12054.	.1410	.712	
error	18	85481.9			
practice	1	1.552E6	10.64	.004	*
mod X practice	1	119660.	.8205	.377	
error	18	145845.			
type X prac	1	140784.	1.954	.179	
modXtypXprac	1	174.05	.0024	.961	
error	18	72045.3			

information ought to be slower.

Two primary considerations must be taken into account, however, which may have played a large role in the actual outcome of the results. First, in this study, equal amounts of hierarchical context are provided in the speech and the pictorial displays. Thus the advantage of context which many previous studies have incorporated only in the speech displays has now also been incorporated in the pictorial displays. It should be noted that the context does not follow the syntactic rules common to the English language; i.e. it is not in "sentence" form. However, the subjects in both groups did go through a training period in which the syntax rules of the experiment were made clear. These rules are the same for both groups, pictorial and speech. It might be argued that the lack of "normal" syntactic structure might have hindered the speech subjects more than the pictorial subjects. A subsequent small study comparing performance with normal syntax and with syntax used in this experiment, at the speech rates used, might help clarify the matter.

The second consideration is that the responses for both types of information display were manual. The model of Figure 1 shows that manual responses are more compatible with spatial input codes than with verbal codes. In this

case, then, compatibility between central processing and responses (C-R compatibility -- see Sandry and Wickens, 1982) seems to override the heavier loading on one encoding/processing channel.

The main effect of Task Type on response time was to be expected. Even though subjects were urged to respond as quickly in the dual task mode as in the single task mode, the allocation of resource capacity to the flying task was a significant drain on the capacities allocated to the emergency response task. Perhaps the most important aspect of the strong significance of Task Type is that the video game does indeed provide the experimenter with a viable "loading" task.

The Modality by Practice interaction on response time also supports the idea that pictorial subjects learn to use the stimulus - response relationships which are not as direct for the speech subjects. Performance of pictorial subjects showed greater improvement with practice than that of the speech subjects. A possible interpretation of this result is that as the subject develops a better mental model of the system, he becomes more confident in his responses, and he makes them more quickly. The subjects receiving verbal information do not enjoy this same advantage, therefore their responses do not speed up with practice as

much as those of the subjects receiving pictorial displays. The fairly direct spatial mapping from the stimulus to the response, a benefit pictures have over speech, may help to strengthen the mental models of the system.

Effects of Task Type and Practice on response accuracy are predictable. As in the response time measurements, the dual task setting demands that attention be allocated away from the emergency responses; thus performance accuracy ought to decrease if the flying task is successfully loading the subject. Also it is natural that the subjects' responses became more accurate with practice. The fact that errors were made suggests that subjects were sufficiently concerned with response time -- they did not always wait to be absolutely sure of their responses before making them.

Based on the error analysis (Appendix C), the largest departure from the expected distribution resulted from speech subjects confusing the three systems (hydraulic, electrical, and propulsion). In the same error class, incorrect system choice, pictorial subjects also deviate from the expected distribution but in the other direction; they make fewer system errors than expected. This supports the idea of spatial advantages in pictorial displays discussed earlier, because there is a direct mapping from the display to the response panel. For example, at the

system level of the display, the hydraulic system is always at the top of the picture. Likewise on the response panel the top two rows of buttons correspond to the hydraulic system. There is no such direct mapping for the speech subjects.

The significant effects of Task Type and Practice on video game scores can receive the same general interpretation as was given for the effects of these factors upon response accuracy. When the subjects were required to concentrate their attention on the game only, their scores were better than when they had to allocate it to the emergencies as well. This indicates that the performance of the video game was resource-limited (see Norman and Bobrow, 1975); i.e. the game was difficult enough to be used as a primary task. Also their scores improved with practice which would be expected.

Experiment One indicated that when the formats of emergency messages are equivalent, i.e. they are both serial in nature with the same amount of context, responses to pictorial messages are faster than to speech messages, especially when the subjects have had practice at the tasks. Referring to the S-C-R model this finding supports the idea that the compatibility between processing and response modes can be more important than distributing the tasks across

different encoding modalities. Also, with practice, subjects with pictorial messages decrease their response times more than subjects with speech messages. These are both important considerations in designing an emergency display system.

Experiment Two

One of the factors which can affect the intelligibility of both speech and pictorial displays is message (speech or CRT update) rate. The original presentation rate was chosen arbitrarily. There is no reason to conclude that that rate is the optimal rate in either modality. This experiment tries to determine the effects of presenting information in the two modes at faster and slower rates at both a low and a high workload situation. This experiment represents a further attempt to understand the trade-offs between pictorial and speech displays which must be considered before implementation of either system.

Method

The method for this experiment was much the same as Experiment One. Again, two tasks were required of the subjects, a tracking video game task and an emergency response task. The task description will not be repeated in this section, but a few differences will be noted.

Since the same subjects and the same equipment was used for this experiment as for Experiment One, no training on either task was required. Subjects executed four single task runs; three single task emergency response runs and one single task video game run. They also were required to "fly" three dual task missions with each mission using a different emergency message rate. The order of the runs was as follows (the order of emergencies re-randomized at each level of rate):

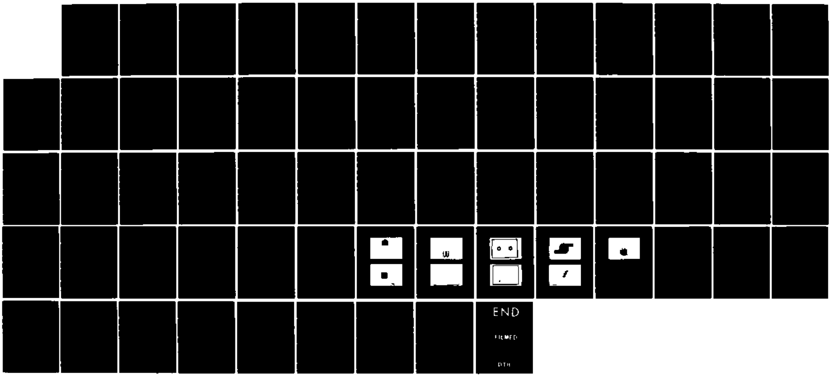
1. Dual Task -- Medium Speed
2. Single Task Emergency Responses -- Medium Speed
3. Single Task Video Game
4. Dual Task -- Fast Speed
5. Single Task Emergency Responses -- Fast Speed
6. Dual Task -- Slow Speed
7. Single Task Emergency Responses -- Slow Speed

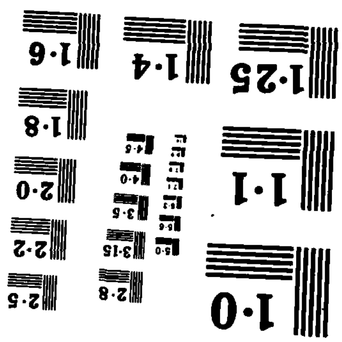
Subjects were given a short rest break following the Single Task Video Game.

The main effects of concern in this experiment included modality, message rate, and task type. For message rate, three different fixed rates were chosen: 1.1 seconds, 1.5 s, and 2.0 s. The "fast" rate, 1.1 s, was limited by hardware; this corresponded to just holding down the slide-advance button on the projector. These rate designation figures correspond to the intervals at which the

AD-A151 917 A COMPARISON OF PICTORIAL AND SPEECH WARNING MESSAGES IN THE MODERN COCKPIT(U) AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH C P ROBINSON DEC 84 2/2

UNCLASSIFIED AFIT/CI/NR-85-23T F/G 9/3 NL





message elements (each of the four phrases or pictures in a message) were initiated. This is illustrated in Figure 10, using the fast rate for the example:

```
Time--------->

     / 1.1s--->/ 1.1s--->/ 1.1s--->/ RT --------> ?

Beep / Element / Element / Element / Element  /  RESPONSE
        One    /  Two    /  Three  /  Four    /
```

Figure 10. Interval Definition of Message Rate

These rates correspond roughly to 110, 80, and 60 words per minute, respectively. In comparison, normal speech rate (reading aloud from printed text) is approximately 145 words per minute.

The experimental design, similar to the first experiment, was a Nested Factorial, with subjects nested under modality. Message Rate and Task Type constituted the factorials. The model used for analysis of variance was is shown in Appendix E.

Results

The pictorial subjects (see Figure 11 and Table 5) responded faster to the emergencies than did the speech subjects ($F(1,18)=9.521$, $p<.006$). Differences in response times (see Figure 12) occurred depending on the presentation rate ($F(2,36)=13.12$, $p<.0001$). A Newman-Keuls test for

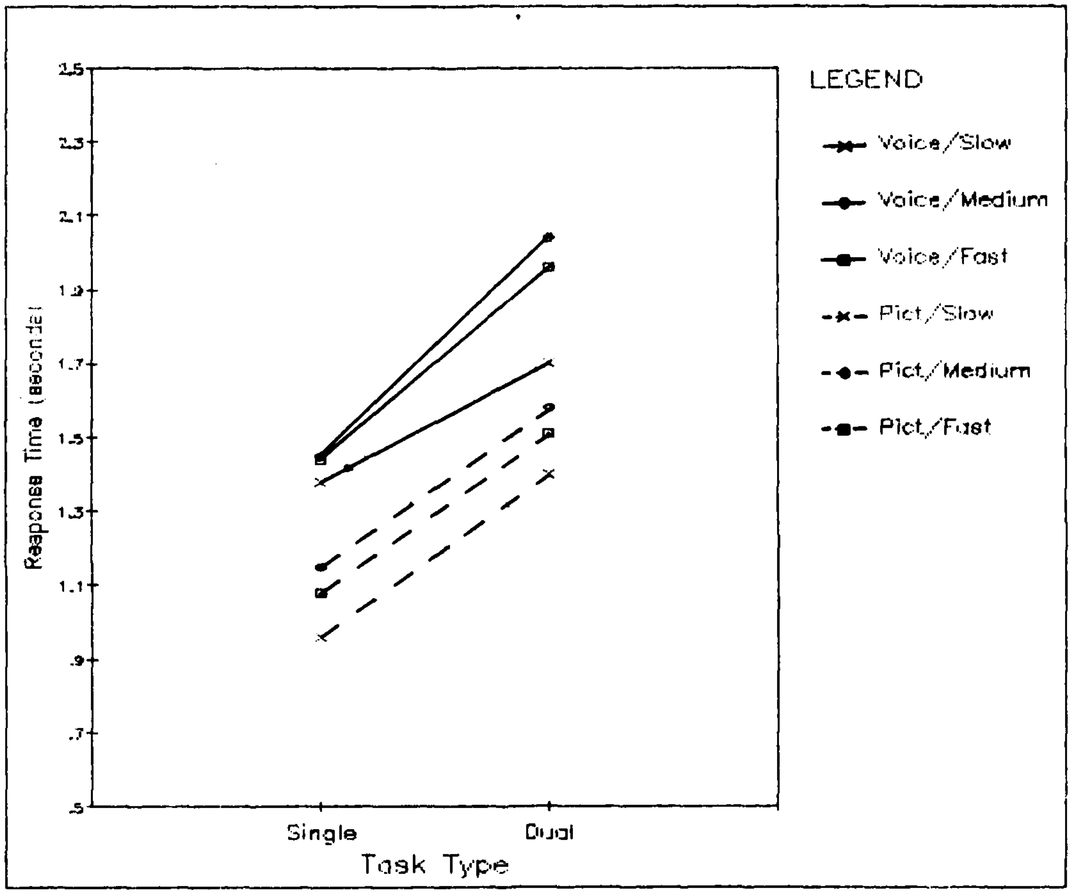


Figure 11. Effects of Modality, Message Rate, and Task Type on Response Time in Experiment Two.

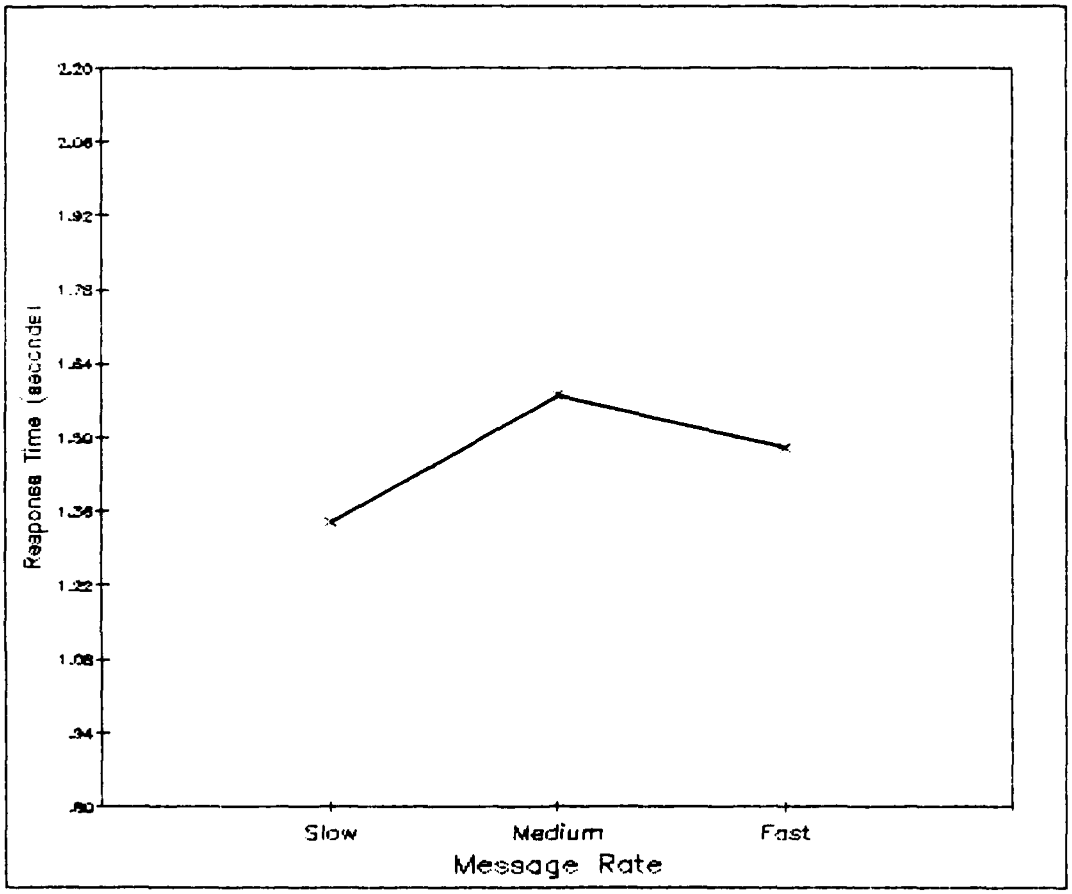


Figure 12. Effect of Message Rate (Collapsed over Modality and Task Type) on Response Time in Experiment Two.

Table 5. Significance Tests for Response Time in Experiment Two.

SOURCE	DOF	MEAN SQUARE	F	p	
residual	0				
mean	1	262.493			
modality	1	4.7760	9.521	.006	*
error	18	.50159			
task type	1	6.9697	69.03	.000	*
mod X type	1	.06816	.6751	.422	
error	18	.10097			
rate	2	.4530	13.12	.000	*
mod X rate	2	.00372	.1077	.898	
error	36	.0345			
type X rate	2	.09484	2.678	.082	
modXtypXrate	2	.11465	3.238	.051	*
error	36	.03541			

Table 6. Significance Tests for Response Accuracy in Experiment Two.

SOURCE	DOF	MEAN SQUARE	F	p	
residual	0				
mean	1	172.800			
modality	1	12.033	4.317	.052	*
error	18	2.787			
task type	1	73.633	26.07	.000	*
mod X type	1	6.533	2.313	.146	
error	18	2.824			
rate	2	4.225	4.379	.020	*
mod X rate	2	.4083	.4232	.658	
error	36	.9648			
type X rate	2	4.008	4.26	.022	*
modXtypXrate	2	.0583	.062	.940	
error	36	.94074			

paired comparisons was performed on the means (Anderson and McLean, 1974). This test showed that response times for the slow message rate were significantly faster than both the medium and the fast rates. The mean response times with the fast rate were shorter than with the medium rate, but this difference was not statistically significant. Responses in the single task situation were faster than in the dual task situation ($F(1,18)=69.03$, $p<.0001$). The three-way interaction of Modality by Task Type by Rate (MTR) was also significant ($F(2,36)=3.238$, $p<.051$). This appears to be due mainly to a smaller degradation in response time between single and dual task runs, at the slow rate, by speech subjects. Effects of other interactions were insignificant at the .05 level.

All three factors had significant main effects on response accuracy (see Figure 13 and Table 6). Speech subjects made marginally fewer errors than the pictorial subjects ($F(1,18)=4.317$, $p<.052$), though accuracy depended on presentation rate ($F(2,36)=4.379$, $p<.020$). A Newman-Keuls test performed on the accuracy means showed that only the errors made in the medium rate were significantly more numerous than those made in the slow rate. The differences in accuracy between fast and slow as well as between fast and medium rates were not significant. More errors were made in the dual task runs than the single

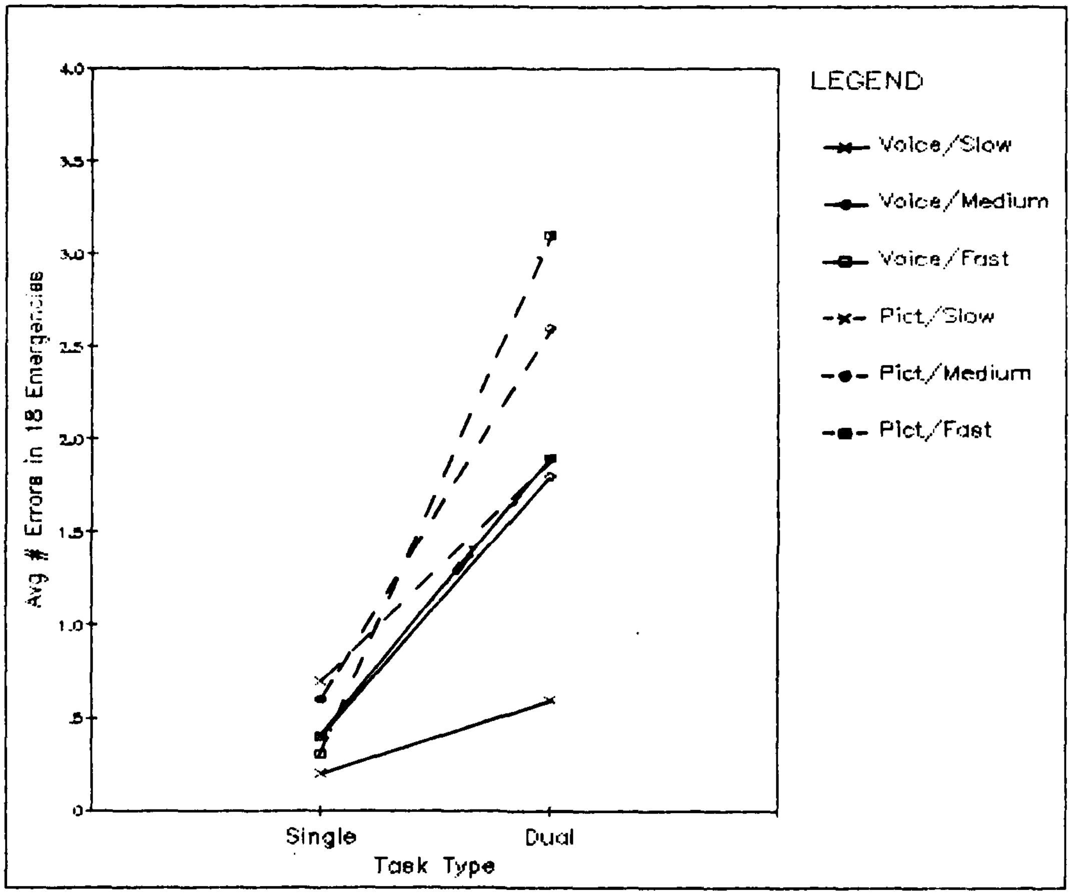


Figure 13. Effects of Modality, Message Rate, and Task Type on Response Accuracy in Experiment Two.

task runs ($F(1,18)=26.07$, $p<.001$). The interaction between task type and presentation rate was significant ($F(2,36)=4.26$, $p<.022$); the degradation at the slow rate in the dual task setting is less than the degradations at the medium and fast rates. The other interactions were not statistically significant. Analysis of the errors (see Appendix C) shows that the distribution of types of errors made by pictorial subjects was significantly different the speech subjects ($X2(3, N=10) = 15.07$, $p<.005$).

The only factor which affected the game scores (Figure 14 and Table 7) in this experiment was Task Type; scores were lower for dual task runs than for single task runs ($F(1,18)=13.78$, $p<.002$). All other main effects and interactions were insignificant.

Discussion

In this experiment, not all measures were affected by message rate. While response time and accuracy were affected, game score was not. This indicates that the subjects followed instruction; during this experiment they protected their primary task. Examination of Figure 12 implies an "inverse U" shaped function within the fixed rates of the experiment, but remember that the mean comparison test showed the fast and medium rates to have essentially the same effect. The relationship with response

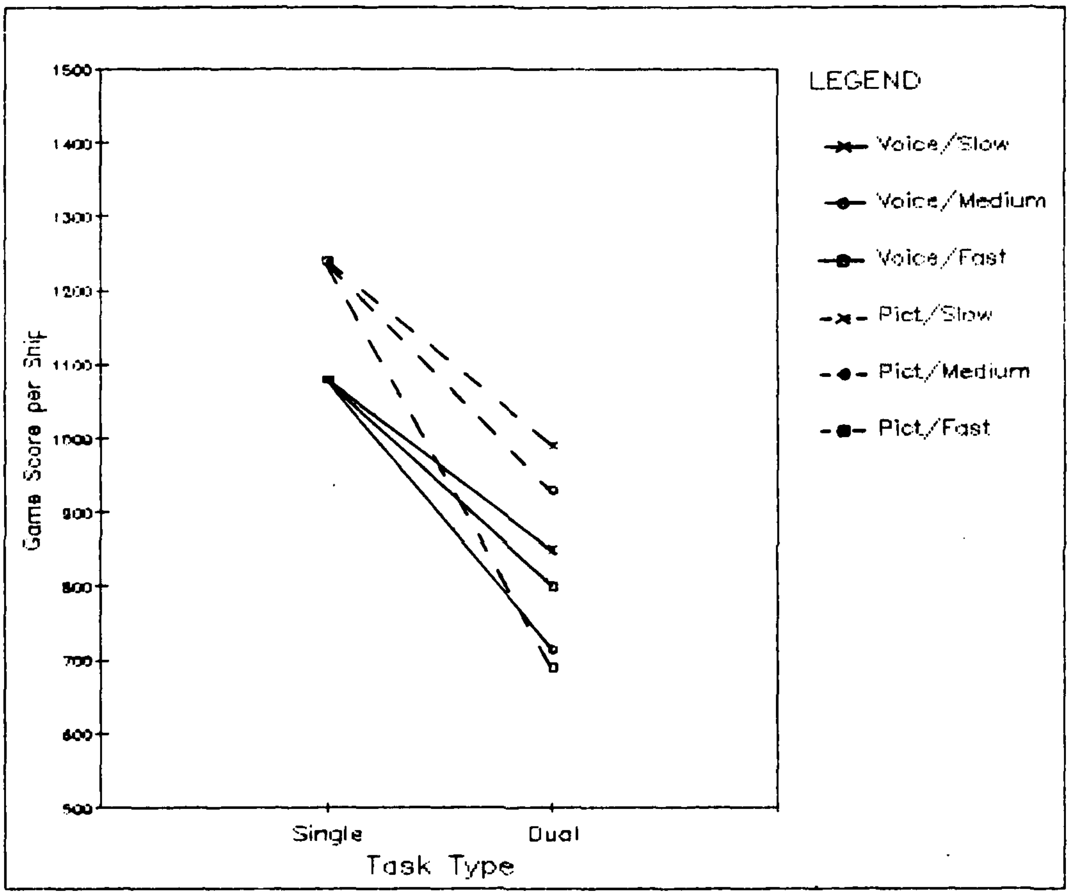


Figure 14. Effects of Modality, Message Rate, and Task Type on Video Game Score in Experiment Two.

Table 7. Significance Tests for Game Scores in Experiment Two.

SOURCE	DOF	MEAN SQUARE	F	p	
residual	0				
mean	1	1.1187E9			
modality	1	441774.	.3662	.553	
error	18	1.2064E6			
task type	1	3.253E6	13.78	.002	*
mod X type	1	46216.8	.1958	.663	
error	18	236005.			
rate	2	75207.	.9743	.387	
mod X rate	2	74969.	.9712	.388	
error	36	77194.			
type X rate	2	75207.	.9743	.387	
modXtypXrate	2	74969.	.9711	.388	
error	36	77194.			

time suggests some "optimal" message rate, a finding similar to one discussed by Simpson and Navarro (1984). They, however, were dealing with higher message rates on the order of 160 words per minute whereas the highest rate in this experiment was 120 words per minute. One factor which could be playing a part here is a confounding of rate with sequence, since the rates were experienced in the order of medium, fast, slow. There may still be a residual learning effect which cannot be separated.

On the other hand, this confounding is not supported by the accuracy measurements, the means of which follow a trend of increasing accuracy with decreasing rate. The interaction between rate and task type seems to be the cause for the finding in the Newman-Keuls test that the difference between fast and slow rate effects were insignificant. In the dual task setting, the accuracy of the fast rate was significantly different from that of the slow rate, as was the medium rate accuracy.

The interaction between rate and modality was not significant, at least not directly. From this information alone, one could conclude that rate had the same effect for pictorial presentation as it did for speech presentation: which does not support the expectation that the pictorial subjects would build a significantly different mental model

of the system than the speech subjects. But, the three-way interaction (MxRxT) effect on response time sheds a different view on the matter. Apparently, in the dual versus single task setting, rate produces greater changes in performance with speech subjects than with pictorial subjects. This finding could be very important in a cockpit environment. When the pilot is in a lower stress environment, performance of the primary task is data-limited, not resource-limited (see Norman and Bobrow, 1975). Addition of a secondary task may not push the limits of the resource pools associated with the two tasks. In a higher stress environment, though, the primary task may transition to a resource-limited process. The secondary task then would probably also be resource-limited.

Between these two scenarios, according to the three-way interaction, different messages and associated rates would incur a variance if the messages were presented by speech that would not be incurred if the messages were pictorial. This variance must be thoroughly understood before implementation of a warning system in a cockpit to minimize potential surprises in future pilot performance.

Finally, the difference between pictorial response times and verbal response times was more significant than in Experiment 1 which again supports the idea that stimulus -

response compatibility available in pictures but not in words may be more important than multiple resources implications discussed previously. By the time this second experiment was completed, the pictorial subjects had more time to learn how to utilize the extra spatial information afforded by the pictures pertaining to the systems and their relationships to the response panel layout. These mappings were probably better developed and more complete than those built by the speech subjects. This possibility becomes more interesting when the main effect of Modality on response accuracy, which approached significance, is considered.

As stated above, the pictorial subjects tended to make more errors than did the speech subjects. Two explanations might be offered for this effect. An analysis of the errors (Appendix C) shows that of the errors made by speech subjects, side reversals and side/both reversals were much less frequent than they were for pictorial subjects. It could be argued from this finding that the verbal transmission of "left" and "right" is more easily processed than a spatial representation. A more probable cause, however, for pictorial subjects' left/right confusions lies in the design of the displays themselves. The standard symbology used in the pictorial displays for this study included coloring the faulty subsystem yellow and placing a yellow "X" over it. For example, if the emergency was in

the left engine, the respective pictorial display would show a green (healthy) engine on the right, and a yellow engine crossed out on the left of the display. On two accounts, subjects' attention was possibly drawn to the right engine. First, while the yellow engine was "lighter" in shade and therefore should have attracted the subject's attention, the green engine may have appeared "brighter" or of higher intensity, overpowering the attractive effect of the lighter color. Secondly, perhaps the "X" caused the subjects to disregard that engine, subconsciously thinking that the "X" meant to look at the other engine, not the crossed-out one. System errors are another large contribution to the differences in the error distributions. Here speech subjects made more errors than would be expected while pictorial subjects made fewer than expected. This indicates the possibility that the spatial aspects of the pictorial messages did provide an important advantage over the speech messages, though in the left/right axis the pictures were not optimized.

The second explanation is one which was also discussed in the Hartzell, et al. (1983) study described in the Recent Cockpit Display Research section of this paper. Theirs was the study of cockpit control and display placement in the modern helicopter, showing that an ipsilateral arrangement was more compatible than a contralateral arrangement. They

found that subjects with ipsilateral controls and displays made more initial movement errors (moving the altitude control in the wrong direction) even though the total response time - including the correction for the initial error - was shorter than that with the contralateral arrangement. As suggested by the authors, this error tendency may have resulted from different strategies employed by the subjects. The subjects with the easier task (ipsilateral condition) tended to initiate the movement, then make corrections. But the subjects with the harder task (contralateral), while sorting out the incompatibility, also thought more about initiating the response in the correct direction. Perhaps a similar process occurred in this study; the speech subjects, while translating from verbal processing to manual/spatial response, spent more time ensuring a correct response.

Experiment Two uncovered some more factors which must be considered in the implementation of a cockpit warning system. Message rate, as well as modality of presentation should be considered. Some situations may be more sensitive to variations in message rate with speech displays than others. If there is a possibility of message rate changing to fit the situation (for example, quicker messages in a time-critical situation), then the designer must be aware of a potential unexpected variance in response to speech

messages. Trade-offs between response speed and accuracy should be considered. What causes them? Can further optimization of the pictorial display design help eliminate them? Can the optimization of pictorial displays be more helpful in supporting an operator's mental model of the system and the stimulus - response relationships? This last question is addressed in the third experiment.

Experiment Three

An assertion previously made is that one of the biggest advantages to spatial pictorial displays is the potential for designing a direct mapping between the display and the response area. As one possibility, the display could even show the exact button to push on a response keyboard. In the cockpit paradigm, the computer would not even have to tell the pilot what the problem is; it could just tell the pilot to push this button or to push that button. Needless to say, this would not be very practical as the pilot needs to feel like she has some control over the airplane. Besides, as long as the final decision to respond or when to respond is going to be left up to the pilot, then she needs to have a reasonable amount of information upon which to base the decision.

Where the advantage does come into play, however, is in the building and maintaining of a sound mental model of the aircraft systems and their interactions. Ideally, everything in the control room should support this mental model and be compatible with it. Not until this condition is met can an optimal performance level be expected. Any information which is presented to the operator should be formatted to be consistent with the model. Any control or response input devices should be designed to maximize compatibility with the model, and therefore with the stimulus information format as well.

The purpose of this third experiment was to see if the two groups of subjects, speech and pictorial, had internalized the displays differently. The internalization to be examined is a spatial mapping of response buttons to the corresponding emergencies. This test was done by comparing performance with response board labels to performance with the labels removed from the keyboard. Given a strong mapping of the system, processing and response performance ought to be superior to performance when these compatibilities are not so complete.

Method

As in Experiments 1 and 2, the subjects were required to perform the two tasks of flying the simulator on an

attack mission in hostile territory, and simultaneously responding to on-board emergency conditions. Two groups of subjects participated, one group receiving generated speech displays and the other receiving pictorial displays. Again, since the same setup and the same subjects were used as in Experiments 1 and 2, a detailed description will not be repeated in this section. For details on equipment, tasks, and subjects, see the method section of Experiment 1.

No special training was required for this experiment since the subjects had already participated in the first two experiments. Six data runs were included in the experiment; two single task (video game), two single task (emergency responses), and two dual task missions. The order of these missions, with the eighteen emergencies randomized at the two levels of "Labels", was as follows:

1. Dual Task -- Labels
2. Single Task, Video Game
3. Single Task, Emergency Responses -- Labels
4. Dual Task -- No Labels
5. Single Task, Emergency Responses -- No Labels
6. Single Task, Video Game

The first three runs of this experiment were the same runs used to collect data for the "Practice" condition of Experiment 1. The message Rate was held constant at 1.5 second intervals, the "medium" rate used in Experiment 2.

The factors of interest in this experiment were Modality, Labels, and Task Type. The design consisted of a Nested Factorial, again with subjects nested under Modality. The factorials therefore were Labels and Task Type. The statistical model for the data analysis was identical to that used in Experiment 1, with Labels substituted for Practice.

Results

In this experiment (see Figure 15 and Table 8) pictorial subjects again responded faster than speech subjects ($F(1,18)=9.523$, $p<.006$), and responses in the single task situations were quicker than in the dual task situations ($F(1,18)=48.90$, $p<.0001$). Responses in the No Label condition were quicker than in the Label condition ($F(1,18)=18.23$, $p<.0005$). In a three-way interaction between Modality, Task Type, and Labels (see Figure 15), speech subjects responded slower with labels than with no labels in dual task runs, but in single task runs they responded at the same speed with or without labels. The pictorial subjects had the same response time difference, when labels were removed, in the dual and the single task runs. However, this three way interaction was not statistically significant ($F(1,18)=3.553$, $p<.076$).

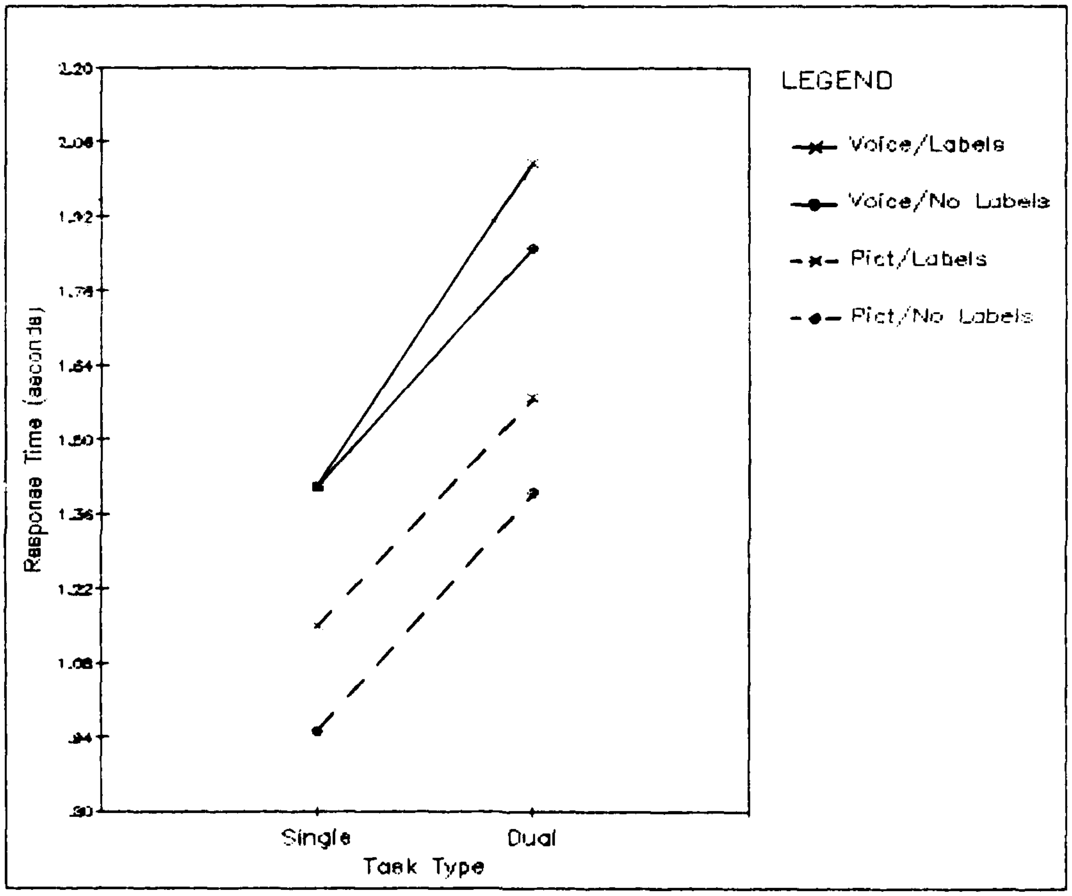


Figure 15. Effects of Modality, Response Panel Labels, and Task Type on Response Time in Experiment Three.

Table 8. Significance Tests for Response Time in Experiment Three.

SOURCE	DOF	MEAN SQUARE	F	p	
residual	0				
mean	1	177.727			
modality	1	3.8194	9.5227	.006	*
error	18	.40108			
task type	1	5.3665	48.90	.000	*
mod X type	1	.12168	1.109	.306	
error	18	.1097			
labels	1	.55778	18.23	.0005	*
mod X labels	1	.01152	.3765	.547	
error	18	.03060			
type X labels	1	.08712	2.480	.133	
modXtypXlabel	1	.12482	3.553	.076	
error	18	.03513			

Table 9. Significance Tests for Response Accuracy in Experiment Three.

SOURCE	DOF	MEAN SQUARE	F	p	
residual	0				
mean	1	180.00			
modality	1	4.050	1.697	.209	
error	18	2.386			
task type	1	76.050	40.08	.000	*
mod X type	1	12.800	6.746	.018	*
error	18	1.8972			
labels	1	1.800	1.111	.306	
mod X label	1	.0500	.0308	.862	
error	18	1.619			
type X label	1	1.250	.7826	.388	
modXtypXlabel	1	5.000	3.130	.094	
error	18	1.597			

An interaction of Modality by Task Type (see Figures 16 and 17, and Table 9) shows that pictorial subjects made fewer errors during single task runs than speech subjects, but in dual task runs pictorial subjects made more errors ($F(1,18)=6.746$, $p<.018$). Also, more errors were made in the dual task missions than the single task runs ($F(1,18)=40.08$, $p<.0005$). An analysis of the errors (see Appendix C) showed no significant difference in the distribution of error types between pictorial and speech subjects.

Video Game scores (see Figure 18 and Table 10) again were higher in single task than in dual task situations ($F(1,18)=36.34$, $p<.0005$). Also, scores were higher with no labels on the response panel than when the labels were present ($F(1,18)=7.628$, $p<.013$). No other factors or interactions were significant.

Discussion

One difficulty encountered in interpreting this data is that the label main effect is confounded with time, so practice may be a significant element of the "label" effect. If this is assumed true, an interesting point comes up when the results of Experiment 1 are taken considered. In that experiment, Practice had a significant effect on response accuracy. If Practice was a main element of the Labels parameter in Experiment 3, then, "Labels" should have at

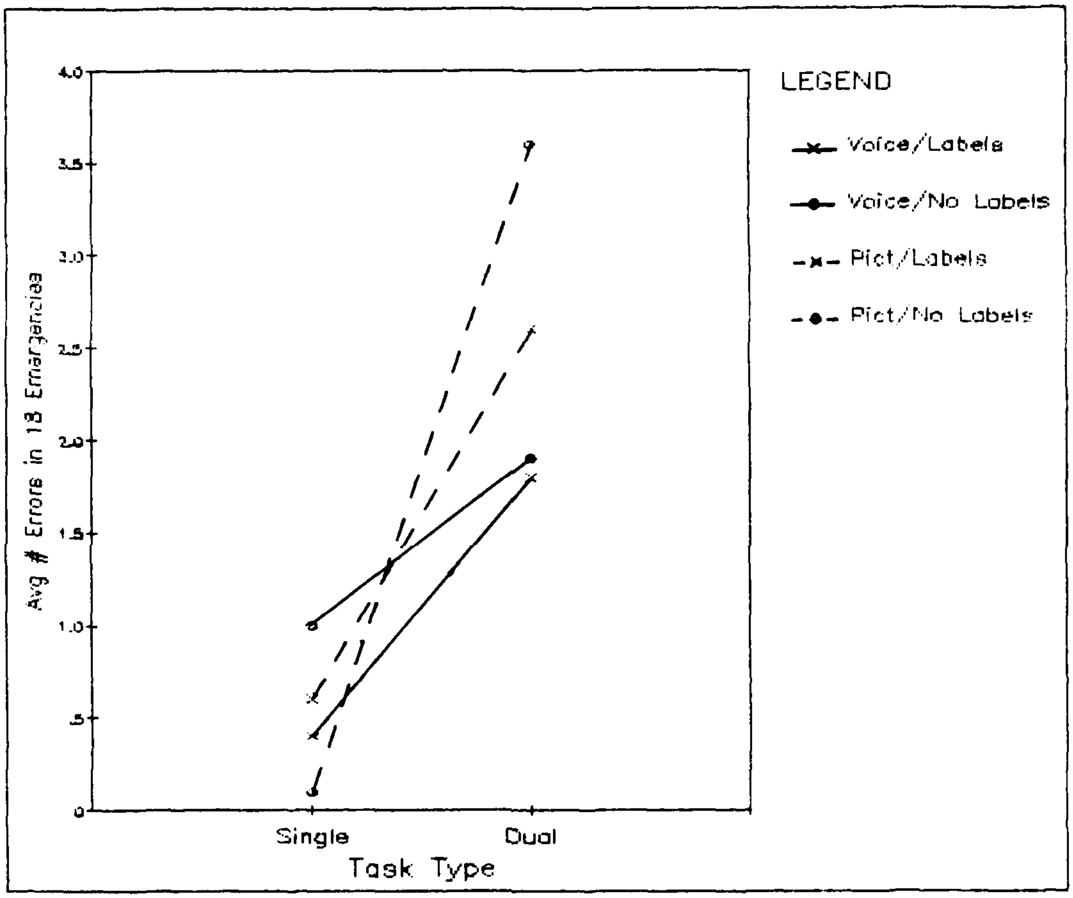


Figure 16. Effects of Modality, Response Panel Labels, and Task Type on Response Accuracy in Experiment Three.

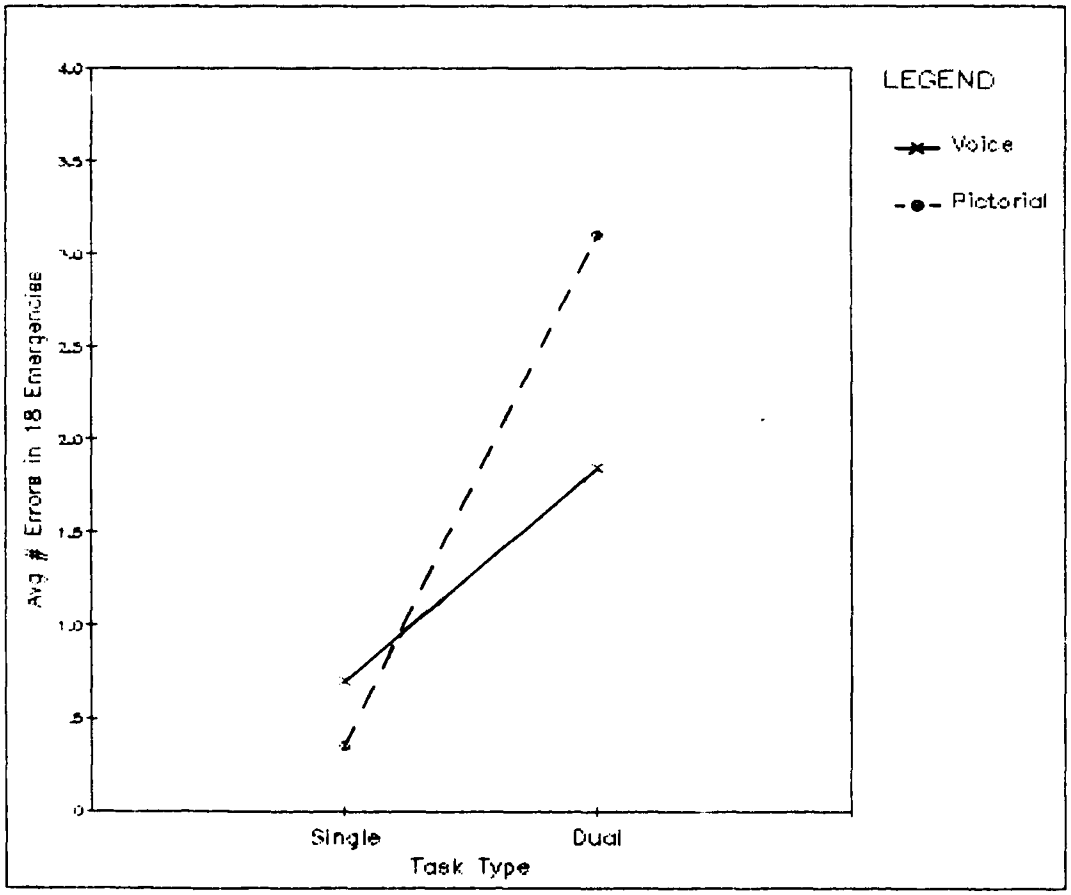


Figure 17. Effect of Interaction between Modality and Task Type on Response Accuracy in Experiment Three.

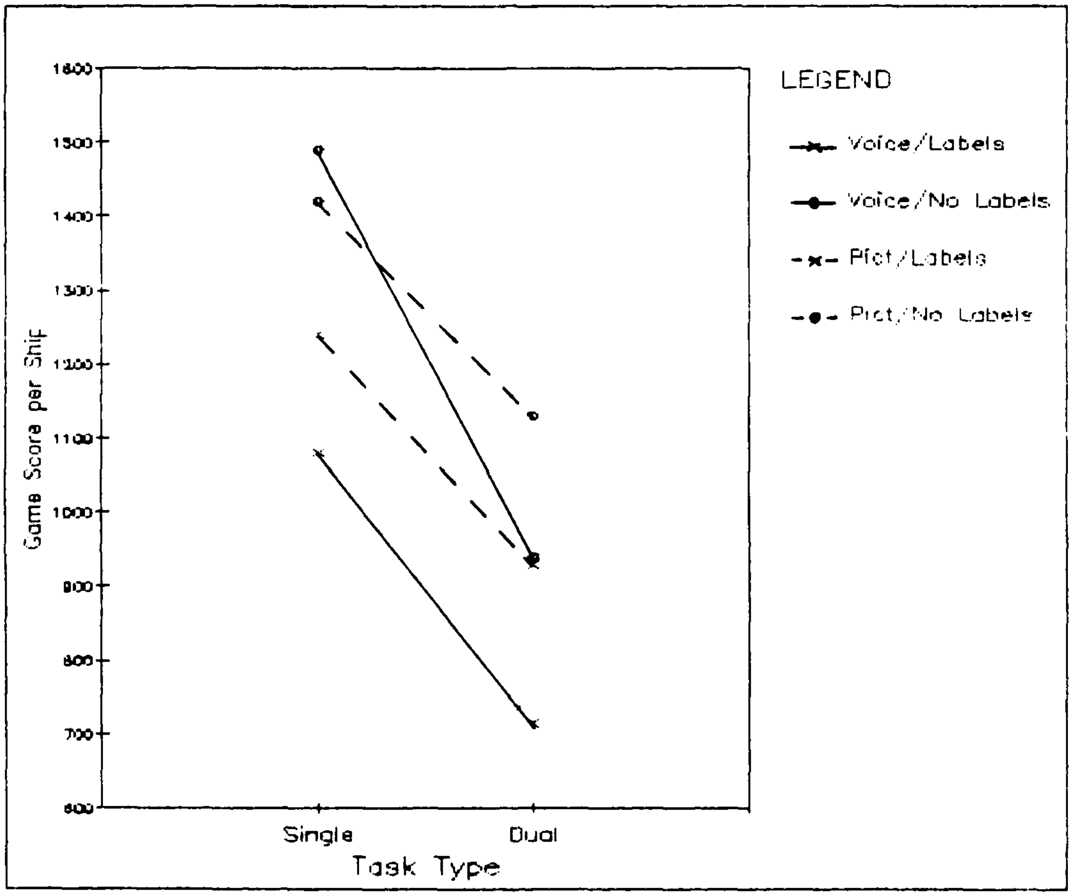


Figure 18. Effects of Modality, Response Panel Labels, and Task Type on Video Game Score in Experiment Three.

Table 10. Significance Tests for Game Score in Experiment Three.

SOURCE	DOF	MEAN SQUARE	F	p	
residual	0				
mean	1	1.1004E9			
modality	1	300737.	.3426	.566	
error	18	877893.			
task type	1	2.8474E6	36.34	.000	*
mod X type	1	122226.	1.559	.228	
error	18	78359			
labels	1	1.3367E6	7.628	.013	*
mod X label	1	85477.	.4878	.494	
error	18	175229.			
type X label	1	33333.	.2153	.648	
modXtypXlabel	1	51359.	.3317	.572	
error	18	154855.			

least approached significance. A possible explanation for the fact that it didn't is that the pure Labels effect, independent of the Practice element, was significant in the opposite direction. In other words, when the labels were removed, there was a degradation in response accuracy, but this effect was cancelled by the Practice effect.

With regards to response time, a trend occurs which is opposite to that evident in the accuracy measure. In Experiment 1, Practice did not have a significant main effect on response time. In Experiment 3, Labels (including any confounding with Practice) did have a significant main effect. But as stated above, responses were quicker without the labels than with them. A possible interpretation for this result is that with the labels available, subjects probably are inclined to read them to be sure that their response decision is correct. When the labels are removed, however, the subjects do not have this luxury; they must simply make the response and hope for the best. In this case there is no excuse to delay because there are no labels to compare their decision with anyhow.

In order for this mode of operation to be successful, i.e. to have a reasonable amount of accuracy along with the decreased response times, the operator must have developed a solid mental mapping or knowledge base of the system and its

interactions with the display information as well as between the displays and the response board. This leads to the question of which type of information best supports the operator's concept of the stimulus-response relationships. If one type was better, a two-way interaction between Modality and Labels could be expected. Alas, this interaction did not appear significant in the experiment, thus the expectation that pictorial subjects would develop better conceptualizations of the S-R relationships was not supported. But when Task Type was added in, the three-way interaction did hint of potential interest. In the single task situation, the speech subjects do not appear to rely on their internalized S-R mappings when the labels are removed, but they do rely on them in the dual task mission. The pictorial subjects rely on their models regardless of whether the task type is single or dual. However, as discussed in Experiment 2, the pictures need to be optimized so that the correlations between the models and response panels are not reversed. The analysis of error type distribution for this experiment (Appendix C) also did not show significant evidence of a difference in mental models or internal representations of the display/response interactions.

The Modality effect on response time, which was marginally significant in Experiment 1 and was significant

in Experiment 2, was once again a strong factor in Experiment 3. The implications of this are the same as those discussed in the other two experiments; the pictorial subjects are more confident of their responses, the spatial compatibility between the pictures and the response panel is greater than the compatibility between the words and the panel. So even though the pictorial subjects use the same modality and code to process the two sets of information while the speech subjects use different resource pools, the pictorial subjects respond more quickly than the speech subjects.

The main effect of Labels on Video Game Score can not be overlooked. Since Practice had a very significant on Score in Experiment 1, one would suspect that it might be the main reason for the "Labels" effect on Score in Experiment 3. Another possibility which might bear further investigation is that when the labels were removed, the subjects took less time and attention away from the various resource pools utilized by the flying task since there were no labels to demand any processing. As a result, since more resource capacity was available for the game task, the game Scores increased in the no label condition.

Experiment 3 did not show any clear difference between the internalizations of the display/response relationships

formed by pictorial and speech subjects. However, the differences in response times did support the possibility that pictorial subjects were able to respond faster than speech subjects because the direct mapping between display and response reduced the number of processing steps.

GENERAL DISCUSSION

In the introduction to this paper, it was suggested that pictorial displays would provide an advantage of stimulus - response compatibility which would not be provided by voice displays. This advantage is a direct mapping between the displays and the response panel which is available in pictorial displays due to their spatial nature. In effect, this may considered as more information being present in the pictures. However, since the tracking task utilizes visual and spatial resource pools, multiple resource theory suggests that secondary information be presented utilizing auditory and verbal resource pools, i.e. voice. The question then arises; is this extra amount of information in the pictorial displays sufficient to overcome the resource advantages of voice displays? The three experiments support in various ways, though not completely, the possibility that the extra spatial information is indeed advantageous over the voice benefits.

Subjects with pictorial displays consistently made quicker responses than subjects with voice displays. This finding suggested that the response decisions were easier to

make; less processing time was required. It also suggests that the pictorial subjects were more confident of their decisions, and thus were able to respond more quickly. The Modality by Practice interaction found in Experiment 1 supported the possibility that the spatial information in pictorial displays helped subjects to learn the response task more quickly than the voice displays did. One potential explanation for this increased learning rate is that subjects with pictorial displays developed mental models of the aircraft subsystem relationships more readily than subjects with voice displays. Since there was a direct mapping between the systems and the response panel, these mental models could be extended to aid in relating the emergency information to the required responses. More likely, however, the quicker response times are a simple result of the extra spatial information in the pictures.

The measure of response accuracy, however, did not entirely support the presumed spatial advantages. In Experiment 2, the main effect of Modality actually favored voice displays, as these subjects made fewer errors than the pictorial subjects. One possible explanation of this is a simple speed-accuracy tradeoff; pictorial subjects respond faster and make more errors as might be expected if the tradeoff didexist. Another possibility is that the pictorial displays were confusing in the left-right

subsystem parameter. In order that this finding not be misconstrued, a further similar study would be recommended, though following a brief study designed to ensure the intelligibility of the pictorial displays. In other words, make sure that the errors made by pictorial subjects are not caused by sub-optimal pictures.

The responses to a questionnaire issued to subjects after completing the three experiments are shown in Appendix D. Comparing questions 4a from the two questionnaires (pictorial and voice) there is a hint that the layout of the pictures was somewhat more confusing than the words. (However, the difference in the mean response levels to this question was not statistically significant.) Based on the analysis of errors shown in Appendix C, the words "left", "both" and "right" were more directive than the pictorial representations of the same. Meanwhile, the same error analysis suggested that voice subjects made many more errors in selecting the "system" than did the pictorial subjects. Referring to the pictures in Appendix A, it can be seen that the system information can be mapped spatially to the keyboard without even translating the system designation into a verbal code. For example, at the system level display, "Hydraulic" is always at the top of the picture, and "Propulsion" is always at the bottom. This corresponds to the keyboard, on which the top two rows of buttons are

dedicated to hydraulic problems, and the bottom two rows are dedicated to propulsion problems. The voice subjects do not have this direct mapping. Comparing the means of the effectiveness ratings in question 4c of the questionnaire (though they were not statistically different), along with the error analysis, there is a suggestion that this direct mapping was indeed helpful to the pictorial subjects.

Assuming the validity of the multiple resources information processing theory, two possible lines of reasoning might have been followed in hypothesizing the results of these experiments. One line would be the following. The primary task is encoded, processed, and responded to using primarily visual, spatial, and manual resource pools. Therefore, the best performance on the secondary task would result from utilizing auditory and verbal pools for encoding and processing, even though the response must be made manually. The other line of reasoning would be that the interference caused by the crossover from verbal central processing to manual response would be enough to outweigh the advantages of having used different resource pools in the first two stages of processing. This second line of reasoning was supported by the three experiments from the standpoint of response time.

In interpreting these results, it is important to remember the possible limit on the generality of their direct application. Neither set of displays, voice nor pictorial, could be considered as optimized in this experiment. The primary intent of this study was not to determine if either display type is better than the other; but to help determine if more research needs to be conducted to find potential advantages of pictorial displays before too many types of alerting systems are delegated to voice displays.

CONCLUSION

With the incorporation of modern computers into today's cockpits, designers are faced with many more options concerning how the pilot and computer may communicate. In particular, two methods of information display are receiving the major focus of research attention. These are computer generated voice and computer generated pictorial displays. In an attempt by the research world to decide which of these methods ought to be used for displaying emergency information, a combination of parametric studies and theoretical arguments have led to use of generated voice. But are all factors being considered? Are the comparisons being made fair comparisons?

In this study, an attempt was made to eliminate some of the advantages that voice has enjoyed in previous studies such as hierarchical context. In this experiment, the messages were formatted so that pictorial messages had the same amount of context as the voice messages. Thus the amount of information requiring processing at each level of the hierarchy was equivalent for pictorial and voice messages. As discussed earlier, this variable has not

always been held constant between the display methods in previous studies.

Also in this study an attempt was made to fully exploit the spatial information which is available in pictures but not directly in words. Often in previous studies there has been no particular correlation between the displays and the required responses, (or stimulus-response compatibility). The results of this study indicate that when this type of compatibility is put into effect, pictorial emergency displays may indeed have advantages over voice displays.

The response method may dictate the information presentation method. Theory states that if the responses to a secondary task can be voice, and if the primary task is visual in nature (as is controlling an airplane), then the secondary information display should be generated voice. The trouble is, that with the current state of technology, voice input systems are limited by their recognition capabilities. The digitized template will not match the pilot's voice input when he is under extreme stress (such as he would be if his engine caught fire over hostile territory) even if he remembered the correct word to input (Williamson and Curry, 1984). Until voice recognition is perfected, manual responses will be preferable for critical inputs.

While information processing theories such as multiple resource and stimulus-central processing-response compatibility theories provide direction for the design of emergency message displays, other concepts must be considered as well. Two of these concepts are the development of mental models, and hierarchical mental organization. Displays should be designed to help develop and support the operator's mental model, or internal representation, of the system and the stimulus/response relationships. If there is a direct mapping from the system to the response board and the displays support this, then the operator will have to go through fewer mental processes (e.g. translating verbal information to spatial response) before making the response. This will in turn reduce the response time, even though two tasks may be drawing from the same resource pools. The displays should also be designed so that at one point in time there is not an overload of information. In past comparisons, pictorial messages have not incorporated hierarchical context which is more inherent in voice displays. When context is provided, the amount of information needing processing at any one time is reduced. In a high workload situation where tasks are resource-limited, this reduction of information is important. This study has provided evidence that when pictorial displays are equated to voice displays in the

amount of context provided, they have certain advantages over the voice displays. These include possible development of a more secure mental model, quicker response times, and better learning characteristics. More consideration of these advantages must be given before implementing too many generated voice displays into the modern control room.

LIST OF REFERENCES

LIST OF REFERENCES

Allport, D.A.; Antonis, B.; and Reynolds, P. On the Division of Attention: A Disproof of the Single Channel Hypothesis. Quarterly Journal of Experimental Psychology, 1972, 24, 225-235.

Anderson, V.L. and McLean, R.A. Design of Experiments: A Realistic Approach. New York: Marcel Dekker, Inc. 1974.

Aretz, A.J. and Calhoun, G.L. Computer Generated Pictorial Stores Management Displays for Fighter Aircraft. Proceedings of the Human Factors Society 26th Annual Meeting, Seattle, Wa., Oct 1982, 455-459.

Bouis, D.; Voss, M.; Geiser, G.; and Haller, R. Visual Versus Auditory Displays for Different Tasks of a Car Driver. Proceedings of the Human Factors Society 23rd Annual Meeting, Boston, Ma., Oct 1979, 35-39.

Bower, G.H.; Clark, C.C.; Lesgold, A.M.; and Winzenz, C. Hierarchical Retrieval Schemes in Recall of Categorized Word Lists. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 323-343.

Broadbent, D.E.; Cooper, P.J.; and Broadbent, M.H.P. A Comparison of Hierarchical and Matrix Retrieval Schemes in Recall. Journal of Experimental Psychology: Human Learning and Memory, 1978, 4(5), 486-497.

Butler, F.; Manaker, E.; and Obert-Thorn, W. Investigation of a Voice Synthesis System for the F-14 Aircraft. Naval Air Development Center, Warminster, PA. TR ACT-81-001. 1981.

Cooper, G.E. A Survey of the Status of and Philosophies Relating to Cockpit Warning Systems. NASA-CR-152071, A03/MF A01 CSCL d5E, Jun, 1977.

Dray, S.M.; Ogden, W.G.; and Vestewig, R.E. Measuring Performance with a Menu-Selection Human-Computer Interface. Proceedings of the Human Factors Society 25th Annual Meeting, Rochester, NY, Oct 1981, 746-748.

Eberts, R.E. Augmented Desplays for Problem Solving. In G. Salvendy (Ed), Advances in Human Factors/Ergonomics: Human-Computer Interaction. Amsterdam: Elsevier, 1984.

Eberts, R.E. and Schneider, W. Computer Assisted Desplays Enabling Internalization and Reduction of Operator Workload in Higher Order Systems, or, Pushing the Barrier of Human Control Beyond Second Order Systems. Proceedings of the Human Factors Society 24th Annual Meeting, Los Angeles, Ca., Oct 1980, 59-62.

Fitts, P.M. and Seeger, C.M. S-R Compatibility: Spatial Characteristics of Stimulus and Response Codes. Journal of Experimental Psychology, 1953, 46, 199-210.

Flight Operating Manual for the F-16-A/B, USAF T.O. 1f-16A-1, April, 1979.

Hakkinien, M.T. and Williges, B.H. Synthesized Voice Warning Messages: Effects of Alerting Cues and Message Environment. Paper Presented at the 26th Annual Meeting of the Human Factors Society, Seattle. Oct 1982.

Hartzell, E.J.; Dunbar, S.L.; Beveridge, R.; and Cortilla, R. Helicopter Pilot Response Latency as a Function of the Spatial Arrangement of Instruments and Controls. DTIC Technical Report ADP001636, Jan 1983.

Hawkins, J.S.; Reising, J.M.; Lizza, G.D.; and Beachy, K.A. Is a Picture Worth 1000 Words -- Written or Spoken? Proceedings of the Human Factors Society 27th Annual Meeting, Norfolk, Va., Oct 1983, 970-972.

Henneman, R.L. and Rouse, W.B. Human Performance in Monitoring and Controlling Hierarchical Large Scale Systems. Proceedings of the Human Factors Society 27th Annual Meeting, Norfolk, Va., Oct 1983, 685-689.

Hollan, J.D. Intelligent Object-Based Graphical Interfaces. In G. Salvendy (Ed), Advances in Human Factors/Ergonomics: Human-Computer Interaction. Amsterdam: Elsevier, 1984.

Hull, C.H. and Nie, N.H. (Eds) SPSS Update 7-9: New Procedures and Facilities for Releases 7-9. New York: McGraw-Hill Book Company, 1981.

Kahneman, D. Attention and Effort. Englewood Cliffs, N.J.: Prentice-Hall, 1973.

Kantowitz, B.H. and Sorkin, R.D. Human Factors: Understanding People-System Relationships. New York: John Wiley and Sons. 1983.

Keppel, G. Design and Analysis: A Researcher's Handbook. Englewood Cliffs, N.J.: Prentice-Hall, Inc. 1973.

Liebelt, L.S.; McDonald, J.E.; Stone, H.K.; and Karat, J. The Effect of Organization on Learning Menu Access. Proceedings of the Human Factors Society 26th Annual Meeting, Seattle, Wa., Oct 1982, 546-550.

Lilliboe, M.L. Evaluation of Astropower Inc. Auditory Information Display Installed in the VA-3B Airplane. NATC TR ST31-22R-63, U.S. Naval Air Station, Patuxent River, MD, 1963.

Mandler, G. Association and Organization: Facts, Fancies, and Theories. In T.R. Dixon and D.L. Horton (Eds), Verbal Behavior and General Behavior Theory. New Jersey: Prentice-Hall. 1968.

Mazza, J.D. A Comparison of Integrated and Conventional Cockpit Warning Systems. Master's Thesis, Naval Postgraduate School, Monterey, CA. Sep 1977.

Miller, D.P. The Depth/Breadth Tradeoff in Hierarchical Computer Menus. Proceedings of the Human Factors Society 25th Annual Meeting, Rochester, NY, Oct 1981, 296-300.

Miller, G.A. The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. The Psychological Review, 1956, 63, 81-97.

Moroze, M.L. and Koonce, J.M. A Comparison of Analog and Digital Scales for Use in Heads-Up Displays. Proceedings of the Human Factors Society 27th Annual Meeting, Norfolk, Va., Oct 1983, 938-940.

Mountford, S.J. and Lea, W.A. Critical Issues in the Development of Speech Technology for Man-Machine Integration. Proceedings of the Human Factors Society 26th Annual Meeting, Seattle, Wa., Oct 1982, 205-206.

Navon, D. and Gopher, D. On the Economy of the Human-Processing System. Psychological Review, 1979, 86, 214-255.

Norman, D.A. Some Observations on Mental Models. In Mental Models, Gentner, D. and Stevens, A. (Eds), Hillsdale, N.J.: Erlbaum. 1982.

Norman, D.A. and Bobrow, D.G. On Data-limited and Resource-limited Processes. Cognitive Psychology, 1975, 7, 44-64.

Patsfall, M. Implications of Dual-Task Performance Variables for Design of Generic Workstations: Literature Review. Naval Biodynamics Laboratory, New Orleans, LA. T.R. 81R014, Nov 1981.

Reising, J.M. Total Cockpit Implications of Electro-Optical Displays. AGARD Lecture Series No. 76 on Electro-Optical Systems, Rome, Olso, and London, 1975.

Reising, J.M.; Emerson, T.J.; and Aretz, A.J. Computer Generated Formats for Advanced Fighter Cockpits. NATO Workshop, Farnborough, UK, 1984, 31.1-31.9.

Reising, J.M. and Kopala, C.J. Cockpit Applications of Computer Graphics. Harvard Computer Graphics Conference, Cambridge, MA, 1982.

Rouse, W.B. Design and Evaluation of Computer-Based Decision Support Systems. In G. Salvendy (Ed.), Advances in Human Factors/Ergonomics: Human-Computer Interaction. Amsterdam: Elsevier. 1984.

Sandry, D. and Wickens, C.D. The Effect of Stimulus - Central Processing - Response Compatibility and Resource Competition on Pilot Performance. Office of Naval Research, Arlington, VA, TR EPL-82-1/ONR-82-1, Apr 1982.

Simpson, C.A. Effects of Linguistic Redundancy on Pilots' Comprehension of Synthesized Speech. NASA TMX 73170, 1976, 294-308.

Simpson, C.A. and Navarro, T. Intelligibility of Computer Generated Speech as a Function of Multiple Factors. Proceedings of the IEEE NAECON, Dayton, OH, 1984, 2, 932-940.

Simpson, C.A. and Williams, D.H. Response Time Effects of Alerting Tone and Context for Synthesized Voice Cockpit Warnings. Human Factors, 1980, 22(3), 319-330.

Stroface, J.F. and Stark, R.T. Operational Test and Evaluation: Verbal Warning System. TAC TR 62-20, Langley Air Force Base, VA. 1963.

Thomas, E.B. Visual Superiority Effect in Memory (Pictures are Worth 1000 Words). Proceedings of the Human Factors Society 27th Annual Meeting, Norfolk, Va., Oct 1983, 714.

Tullis, T.S. An Evaluation of Alphanumeric, Graphic, and Color Information Displays. Human Factors, 1981, 23(5), 541-550.

Werkowitz, E. Ergonomic Considerations for the Cockpit Applications of Speech Generation Technology. Proceedings of the Voice Interactive Systems Society: Applications and Payoffs, May 1980, 295-309.

Wickens, C.D. The Structure of Processing Resources. In R. Nickerson and R. Pew (Eds), Attention and Performance VIII. New York: Erlbaum. 1980.

Wickens, C.D. Engineering Psychology and Human Performance. Columbus, OH: Charles E. Merrill. 1984.

Wickens, C.D.; Sandry, D.L.; and Vidulich, M. Compatibility and Resource Competition Between Modalities of Input, Central Processing, and Output. Human Factors, 1983, 25(2), 248-277.

Wickens, C.D.; Vidulich, M.; Sandry, D.L.; and Schiflett, S. Factors Influencing the Performance Advantage of Speech Technology. Proceedings of the Human Factors Society 25th Annual Meeting, Rochester, NY, Oct 1981, 705-709.

Wicker, J.E. Some Human Factors Aspects of a Real-Time Voice-Interactive System in the Single-Seat Fighter Aircraft. Proceedings of the Voice Interactive Systems Society: Applications and Payoffs, May 1980, 265-275.

Williamson, D. and Curry, D. Voice I/O Effectiveness Under Heavy Task Loading. Journal of the American Voice I/O Society, May 1984.

APPENDICES

Appendix A. Sample Pictorial Displays

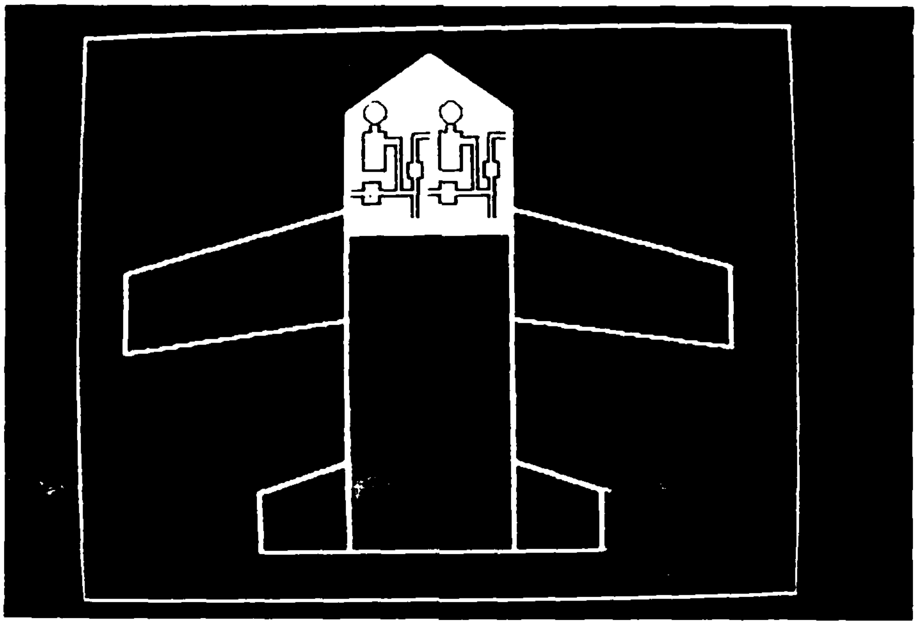


Figure 1A. Hydraulic System

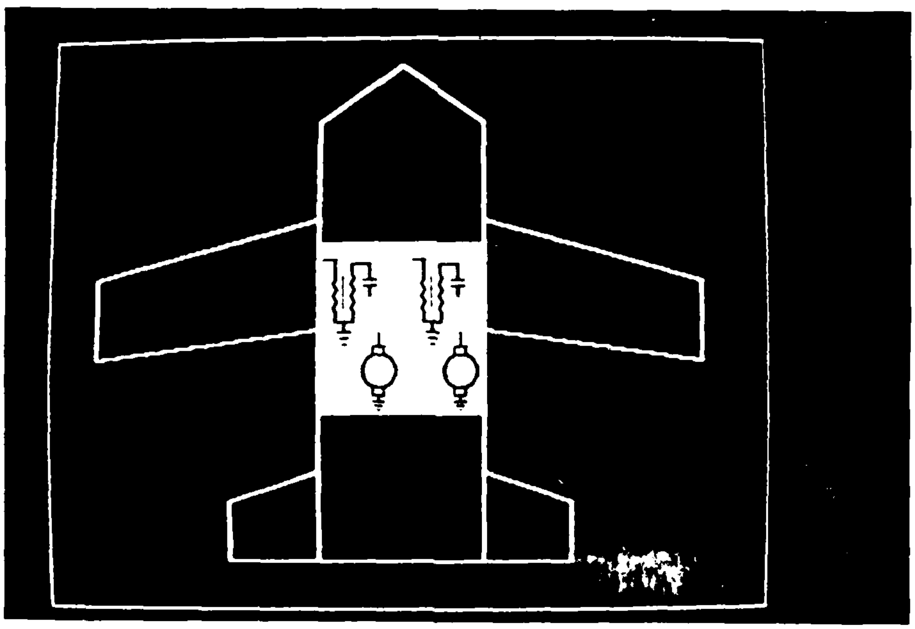


Figure 2A. Electrical System

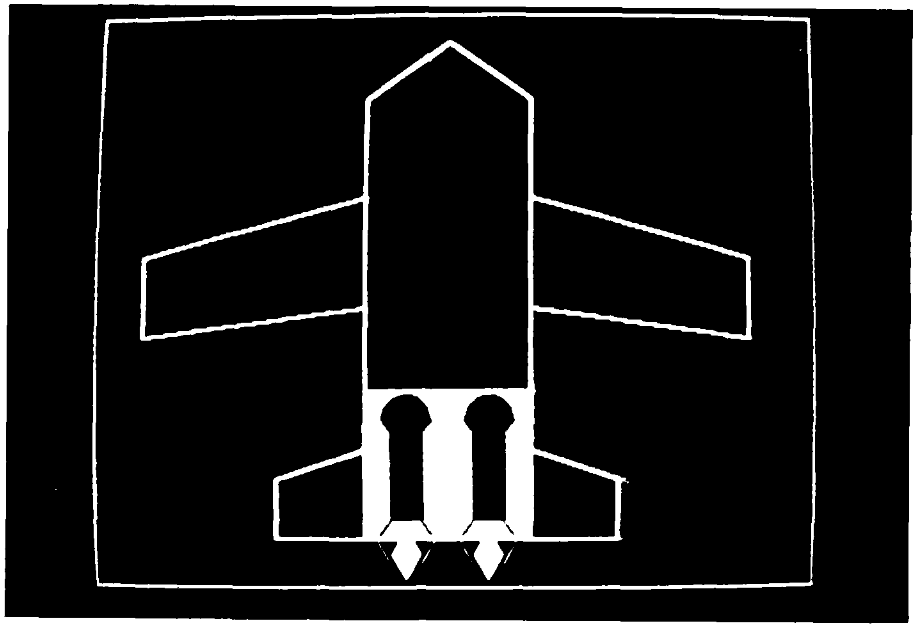


Figure 3A. Propulsion System

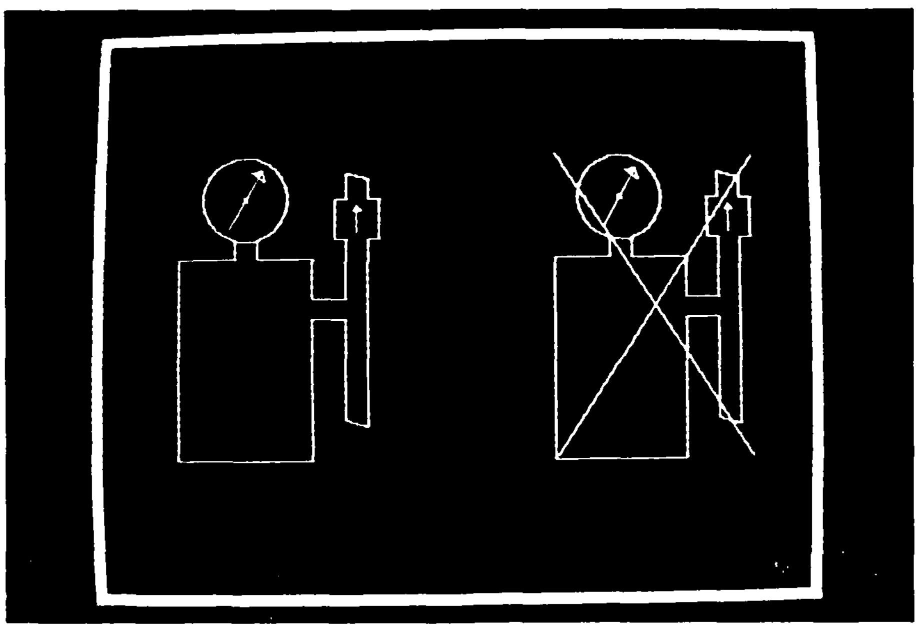


Figure 4A. Right Pumpline

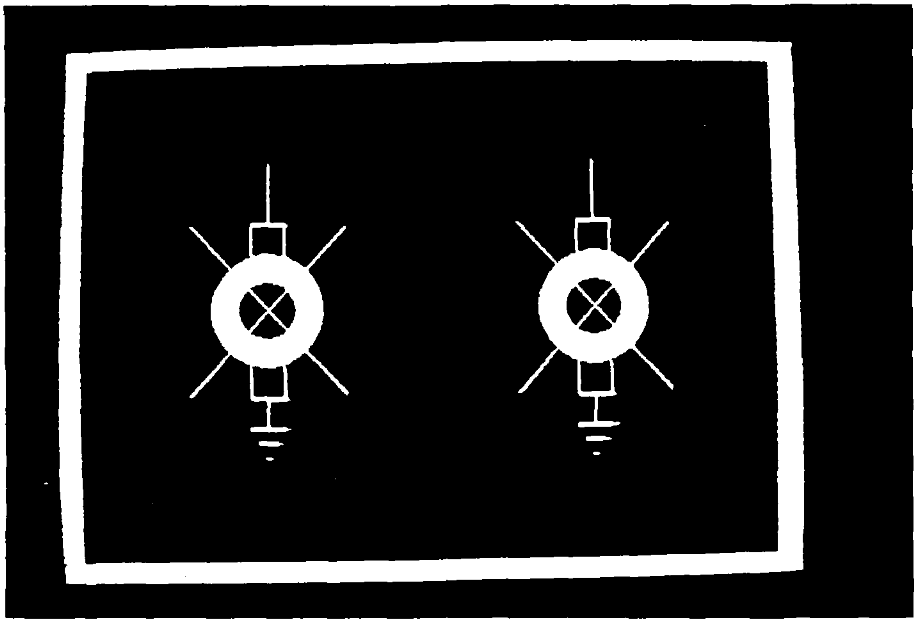


Figure 5A. Both Generators

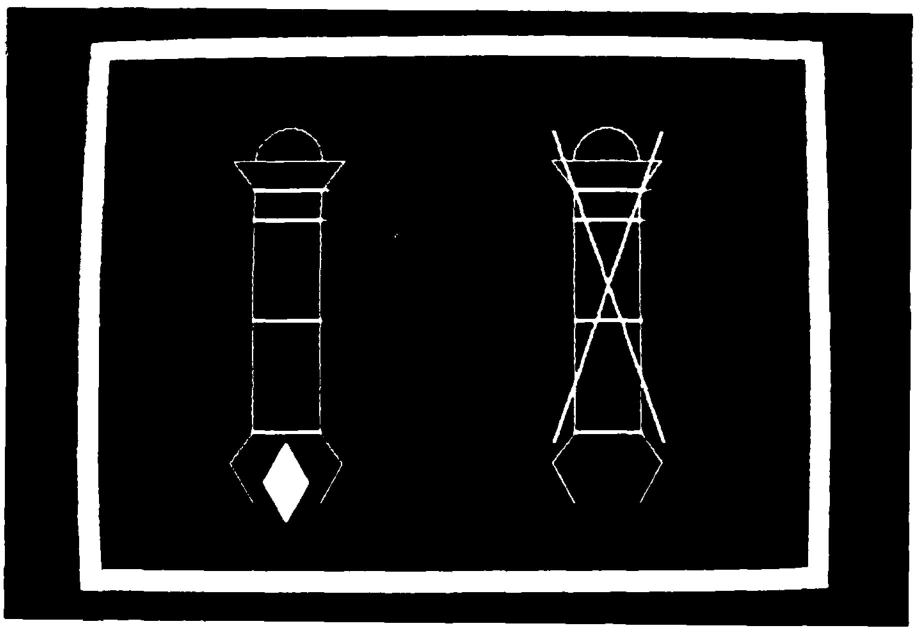


Figure 6A. Right Engine

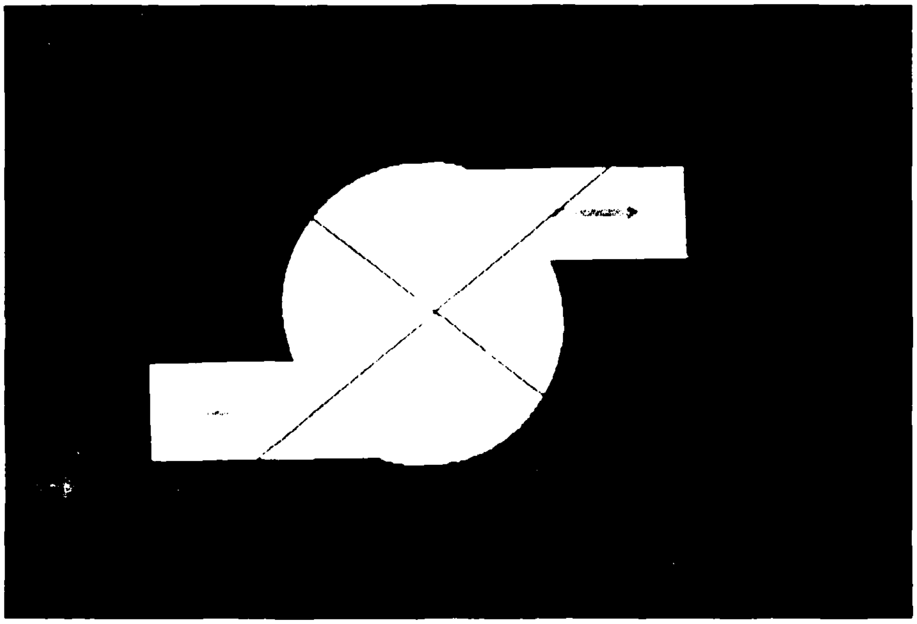


Figure 7A. Pump Fail

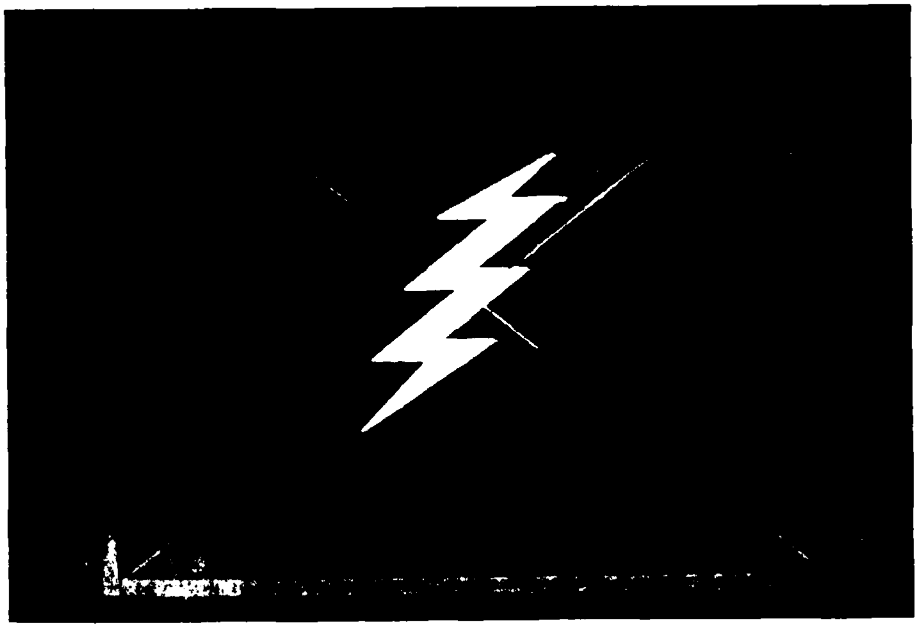


Figure 8A. Power Out

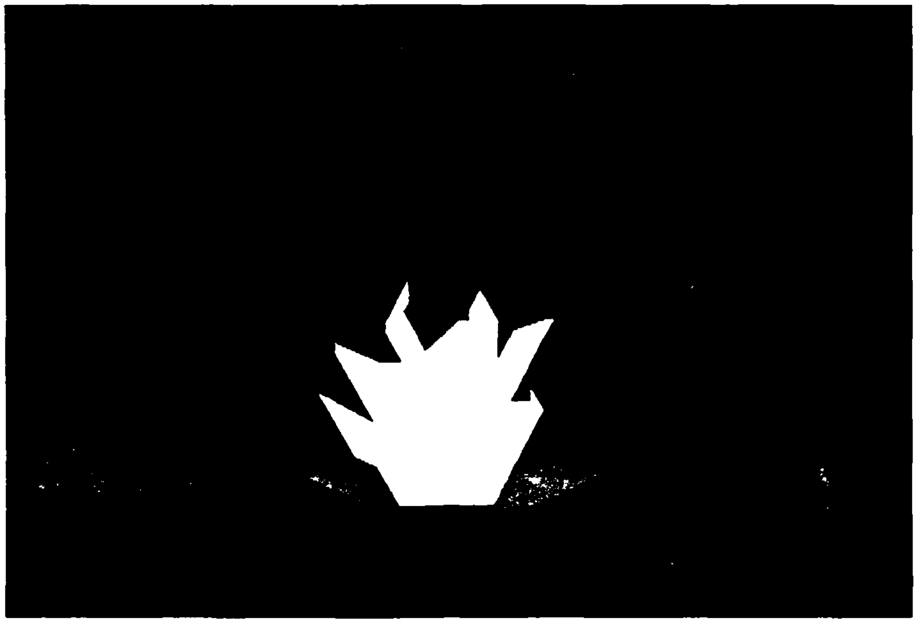


Figure 9A. Fire

Appendix B. Training Scripts

"The purpose of this experiment is to find out the effects of different methods of presenting information to you while you are busy paying attention to something else. What you will be doing is playing an electronic game (Coleco-Vision) in which you fly an aircraft through hostile territory. Your mission (should you decide to accept it) is to knock out as many of the enemy's systems as possible. Meanwhile, you must keep yourself alive because the enemy will be trying to knock out as much of you as possible! While you are flying along, however, you will have problems with your own aircraft. It is a now model, and the bugs have not yet been completely worked out. For example, one of your engines might catch fire, a hydraulic pump may break, or a generator may fail. When something like this happens, you must respond to the emergency as quickly as possible, i.e. before your plane explodes or you lose control of it due to a system failure.

When an emergency occurs within your aircraft, the onboard computer will analyze the problem and notify you, so that you can decide what to do. First you will hear an alarm. Following the alarm, a series of four pictorial displays will flash up on this screen. For example, you will see a "warning signal" followed by a picture which shows you what part of the plane the problem is in

(electrical, powerplant, or hydraulic), a picture showing which part of the system has a problem (left, right, or both), and finally a picture showing you exactly what the problem is (e.g. fire, broken pump, generator not putting out full power).

There are eighteen possible emergencies which can occur in your plane. We will go through them shortly. When one of these occurs, you must respond as quickly as possible by pushing the appropriate button on this keyboard. As you see, you have six rows and three columns of buttons to use. That makes eighteen buttons, which is how many emergencies there are. Each emergency has its own button. For example, if you had an engine fire, when you hit the correct button you might activate the fire extinguisher before your plane blows up.

What we'll do first is let you play the game for about 20-30 minutes so you can get used to it. You won't have to worry about any emergencies cropping up, just play and have fun. The stick in your right hand controls the altitude of your plane, and the one in your left hand is the throttle: it controls your forward speed. You can shoot forward with this trigger and drop bombs with this button.

****

Take single task (game) measure after 25 minutes.

****

Emergency Training

Now that you've had some fun, we're going to make the game even more fun. As I said before, while you are flying, certain things will go wrong with your plane. I'm going to teach you what the different problems can be. Many of the emergencies are related to each other, so that will help you remember them. Also remember that when the emergency occurs, the computer on board your ship will tell you exactly what the problem is; all you have to do is respond as quickly as possible to correct the problem before it is too late.

****

Flip through the demo slides while giving this instruction, and point out the correct response buttons.

****

There are three different systems in your aircraft that might give you problems. They are the HYDRAULIC SYSTEM, the ELECTRICAL SYSTEM, and the PROPULSION SYSTEM.

The hydraulic system is what gives you control of your ailerons, rudder, flaps, etc, (your directional controls). If you lose your hydraulic system, you lose control of your craft. There are two main parts of the hydraulic system, the left pumpline and the right pumpline. Either one or both of the lines can malfunction. Thus, following the hydraulic picture you may see a picture depicting problems in the LEFT PUMPLINE, BOTH PUMPLINES, or the RIGHT PUMPLINE. Two things can happen to them. One, you can have a PUMPFAIL. This is a critical problem because it means that your hydraulic system is useless: you have lost control. You can save yourself, however, by immediately pushing the right one of these buttons which will engage the backup system. The other problem you might have is LOW PRESSURE in the pumplines. This is a dangerous situation which will escalate if you don't respond immediately with one of these buttons."

This explanation continued, to cover the electrical and propulsion systems, in the same fashion. At each underlined word, the subject was shown the corresponding picture.

With the voice subjects, the same training procedure was followed, but instead of showing pictures, the digitized words were played.

Appendix C. Chi-Square Tests for Error Type Distributions

Table 1C. Chi-Square Tests for Error Type Distributions

Response errors were broken down into four classifications:

1. Left/Right Reversal
2. Type of Emergency within Subsystem
3. Incorrect System Choice
4. Left or Right reversed with Both

Experiment One

	Pictorial				Voice				
Class	f	F	X2	p	f	F	X2	p	sum
1	11	11	0	.112	8	8	0	.110	19
2	40	40	0	.408	30	30	0	.411	70
3	10	18	3.56	.102	21	13	4.92	.288	31
4	37	29	2.21	.378	14	22	2.91	.192	51
Total:	98	98	5.77		73	73	7.83		171

TOTAL X2(3, N=10) = 13.60, p<.005

Experiment Two

	Pictorial				Voice				
Class	f	F	X2	p	f	F	X2	p	sum
1	17	13	1.23	.157	4	8	2.0	.066	21
2	30	37	1.32	.278	28	21	2.33	.459	58
3	17	22	1.14	.157	17	12	2.08	.279	34
4	44	36	1.78	.407	12	20	3.2	.197	56
Total:	108	108	5.47		61	61	9.61		169

TOTAL X2(3, N=10) = 15.08, p<.005

Experiment Three

	Pictorial				Voice				
Class	f	F	X2	p	f	F	X2	p	sum
1	9	8	.13	.095	4	5	.2	.071	13
2	37	40	.23	.389	26	23	.39	.464	63
3	19	21	.19	.2	15	13	.31	.268	34
4	30	26	.62	.316	11	15	1.07	.196	41
Total:	95	95	1.17		56	56	1.97		151

TOTAL X2(3, N=10) = 3.14, p<.500

Appendix D. Cummulative Questionnaires

CUMULATIVE QUESTIONNAIRE (for PICTORIAL subjects)

1. How difficult did you find concentrating on the two tasks ("flying", and responding to emergencies) simultaneously?

very easy						very difficult
1	2	3	4	5	6	7
			1	5	3	1

2. After how many slides were you able to determine what each emergency was?

1 slide	2 slides	3 slides	4 slides
		4	6

3. Did the slides follow a logical order in identifying each emergency?

No, chaos.					Yes, logical order	
-3	-2	-1	0	1	2	3
					4	6

4. Please indicate how helpful each of the following was in aiding your responses?

	Very Distracting						Very Helpful
4a) Layout of the Pictures:	-3	-2	-1	0	1	2	3
				1	2	5	2
4b) Sequence of Pictures:	-3	-2	-1	0	1	2	3
					4	1	5
4c) Layout of the Keyboard:	-3	-2	-1	0	1	2	3
		1		2	2	4	1
4d) Presence of Labels:	-3	-2	-1	0	1	2	3
				2	2	5	1
4e) Format of Labels:	-3	-2	-1	0	1	2	3
			1	1	3	4	1

5. Are you a licensed, but non-military pilot?
YES 2 NO 8

6. Which of the three message speeds was best for you?

a. None of them; I would have preferred them slower.

1 b. The slowest of the three that I tried.

9 c. The middle speed I tried.

d. The fastest speed I tried.

e. None of them; I would have preferred them faster.

CUMULATIVE QUESTIONNAIRE (for VOICE subjects)

1. How difficult did you find concentrating on the two tasks ("flying", and responding to emergencies) simultaneously?

very easy					very difficult	
1	2	3	4	5	6	7
	1	1	2	5		1

2. In each emergency, you were given four words (or two-word phrases) to describe the problem. After how many words/phrases were you able to determine what each emergency was?

1 phrase	2 phrases	3 phrases	4 phrases
		2	8

3. Did the phrases follow a logical order in identifying each emergency?

No, chaos.					Yes, logical order	
-3	-2	-1	0	1	2	3
1					2	7

4. Please indicate how helpful each of the following was in aiding your responses?

	Very Distracting						Very Helpful
4a) Directional attributes: of the words.	-3	-2	-1	0	1	2	3
					2	2	6
4b) Sequence of Phrases :	-3	-2	-1	0	1	2	3
			1	1	1	3	4
4c) Layout of the Keyboard:	-3	-2	-1	0	1	2	3
		3	2		2	2	1
4d) Presence of Labels:	-3	-2	-1	0	1	2	3
			2	1	5	1	1
4e) Format of Labels:	-3	-2	-1	0	1	2	3
				3	1	5	1

5. Are you a licensed, but non-military pilot?
YES 1 NO 9

6. Which of the three message speeds was best for you?

a. None of them; I would have preferred them slower.

4 b. The slowest of the three that I tried.

1 c. The middle speed I tried.

4 d. The fastest speed I tried.

1 e. None of them; I would have preferred them faster.

Appendix E. Statistical Models

Model for Analysis of Variance in Experiment One

$$Y_{ijkl} = u + M_i + S(i)_j + x(ij) + T_k + MT_{ik} + ST(i)_{jk} + w(ij) + P_l + MP_{il} + SP(i)_{jl} + TP_{kl} + MTP_{ikl} + STP(i)_{jkl} + e(ijkl)$$

where

- Y_{ijkl} = response time, accuracy, or game score
- u = overall mean
- M_i = effect of Modality, $i=1-2$
- $S(i)_j$ = effect of Subject within Modality, $j=1-10$
- $x(ij)$ = restriction error caused by restriction on randomization of task type
- T_k = effect of Task Type, $k=1-2$
- MT_{ik} = interaction of Modality and Task Type
- $ST(i)_{jk}$ = interaction of Subject within Modality and Task Type
- $w(ij)$ = randomization restriction error
- P_l = effect of Practice, $l=1-2$
- MP_{il} = interaction of Modality and Practice
- $SP(i)_{jl}$ = interaction of Subject within Modality and Practice
- TP_{kl} = interaction of Task Type and Practice
- MTP_{ikl} = Three-way interaction between Modality, Task Type, and Practice
- $STP(i)_{jkl}$ = Three-way interaction between Subject within Modality, Task Type, and Practice
- $e(ijkl)$ = error term

Model for Analysis of Variance in Experiment Two

$$Y_{ijkl} = u + M_i + S(i)_j + x(ij) + T_k + MT_{ik} + ST(i)_{jk} + w(ij) + R_l + MR_{il} + SR(i)_{jl} + TR_{kl} + MTR_{ikl} + STR(i)_{jkl} + e(ijkl)$$

where

- Y_{ijkl} = response time, accuracy, or game score
- u = overall mean
- M_i = effect of Modality, $i=1-2$
- $S(i)_j$ = effect of Subject within Modality, $j=1-10$
- $x(ij)$ = restriction error caused by restriction on randomization of task type
- T_k = effect of Task Type, $k=1-2$
- MT_{ik} = interaction of Modality and Task Type
- $ST(i)_{jk}$ = interaction of Subject within Modality and Task Type
- $w(ij)$ = randomization restriction error
- R_l = effect of message Rate, $l=1-3$
- MR_{il} = interaction of Modality and Rate
- $SR(i)_{jl}$ = interaction of Subject within Modality and Rate
- TR_{kl} = interaction of Task Type and Rate
- MTR_{ikl} = Three-way interaction between Modality, Task Type, and Rate
- $STR(i)_{jkl}$ = Three-way interaction between Subject within Modality, Task Type, and Rate
- $e(ijkl)$ = error term

Model for Analysis of Variance in Experiment Three

The same model was used as for Experiment One, except that Labels was substituted for Practice.

END

FILMED

5-85

DTIC